UNIVERSITY OF LINCOLN

Assessing the predictive value of the UK-CDI for early identification of developmental language delay

Janine Just

A thesis submitted in partial fulfilment of the requirements of the University of Lincoln for the degree of Doctor of Philosophy

May 2018

Acknowledgements

First of all, I would like to thank my supervisors Prof. Kerstin Meints, Dr. Jon Slack and Prof. Caroline Rowland for all their patience, support and knowledge. I would also like to thank Prof. Graham Law who shared his statistical knowledge with me.

I would also like to thank all the parents and children who offered to take part in my longitudinal research. This research would have not been possible without their continuous participation.

Furthermore, I would like to thank my colleagues Mirena Dimolareva and Victoria Brelsford at Lincoln Infant and Child Development Lab for their emotional support and the funny times we had together.

I would also like to thank my beautiful friends in Lincoln, Germany and France for listening to me and supporting me with their kind words.

Finally, my deepest gratitude goes to Paul, my mother, father, and grandparents for their encouragement to complete the PhD, for believing in me and always being there for me.

Table of Contents

List of Tables

Abstract

The UK-Communicative Development Inventory (CDI) is an adaptation of the MacArthur-Bates CDI questionnaires and has been newly developed, standardised and normed for British children between 8 – 18 months. This parent-report instrument assesses children's communication and language. Research in other languages has shown that CDI instruments assessing infant language show good stability with language up to the preschool years on a group level. However, the prediction of language delays/disorders for individual children was unsatisfactory.

This research examined the predictive validity of the UK-CDI for the first time. The aim was to establish if the UK-CDI subscales (Gestures, Phrases Understood, Production and Comprehension) were associated with later language scores up to 36 months, if continuity of language depended on UK-CDI ability group (low, low-average, average-high, high), if other factors needed to be considered to predict later language, and if language delay at 24 or 36 months could be predicted for individual children using the UK-CDI at 12 or 18 months.

Families were from the East Midlands (UK) and took part at four time points (N = 82). Parents completed the UK-CDI and Family Questionnaire at 12 and 18 months, the Lincoln Toddler CDI at 24 months and the 3-year parent report language measure and the British Ages and Stages communication subscale at 36 months. At 18 and 24 months, children also participated in standardised tests assessing language (Preschool Language Scales) and, cognitive and motor ability (Bayley Scales of Infant and Toddler Development).

Results showed that UK-CDI scores at 12 and 18 months were correlated with language scores up to 36 months, albeit language was more stable from 18 months onwards. The associations were usually strongest with the same category at the closest follow-up testing. In addition, stronger correlations were found between UK-CDI scores and other parent-report measures compared to in-person assessments. When investigating the stability of language separately for ability group, language was most stable for high ability children. Furthermore, high ability children remained significantly better than low ability children on most language measures up to 36 months. However, low ability children improved over time in terms of vocabulary but continued to show slow grammar development. In addition, early language (UK-CDI scores) was the best predictor of future language. The other factors (i.e., prematurity, gender, SES, sibling status, ear infections, sleep, family risk of dyslexia or speech or language problems, cognitive and motor skills) influenced later language but their contributions were not consistent and depended on the outcome measure used and the time tested. Therefore, only UK-CDI scores at 12 and 18 months were used to predict language delay at 24 and 36 months using receiver operating characteristic curves. To achieve clinically useful levels of classification accuracy, the UK-CDI cut-off scores had to be higher than the norm-referenced 25th percentile. Depending on the criteria used for delay, Production and Gestures at 18 months predicted delay at 24 months and Production at 18 months predicted delay at 36 months. Implications of using high cut-off scores were discussed.

1. Introduction

This research assesses the predictive validity of the newly adapted UK CDI: Words and Gestures (UK-CDI:WG). This parent-report questionnaire was adapted into British English, standardised and normed for UK children aged between 8 and 18 months, as part of a fully-funded ESRC (Economic and Social Research Council) project (Alcock, Meints, & Rowland, 2017; Alcock et al., in prep). The CDI consists of a word checklist about children's receptive and expressive vocabulary, phrases children understand and gestures they use. Thus, the UK-CDI is able to describe children's early communication status. As part of the above project, the UK-CDI was investigated for its internal consistency, test-retest reliability and concurrent validity.

Until now, there has been no information provided on the predictive validity of the UK-CDI:WG. Therefore, this PhD investigates the predictive validity of this newly developed instrument for young UK children between 12 and 36 months of age. Children's language and cognitive development was tracked in 82 children from the East Midlands region of England, UK. They were first tested aged around 12 months and followed up at 18, 24 and 36 months.

In order to introduce the reader to the predictive value of CDIs, the first chapter summarises the literature on early language development and explains factors which put children at elevated risk for language delay. The second chapter describes the purpose of CDIs and critically assesses the measure. Chapter three reviews the literature on the predictive validity of CDI:WGs. This leads to the next chapter about the current study which outlines its importance, the aims of this project and the research hypotheses. Chapter five introduces the reader to the data collected in this study explaining participant characteristics, descriptive statistics and distribution of language data as well as possible gender differences in terms of language scores. Furthermore, chapter five examines and discusses the relationships between the UK-CDI and later UK-CDI and other language scores. Chapter six investigates the language stability over time of UK-CDI ability groups. Different hypotheses are being tested and results are being discussed. Chapter seven assesses and discusses the contributions of family, biological and environmental factors as well as early communication on later language development. Chapter eight uses individual-level analyses to investigate if the UK-CDI can be used at 12 or 18 months to predict language delay in individual children at 24 or 36 months. The results are being discussed. Chapter nine forms the conclusion which summarises the findings and points out future directions for research.

1.1. Early language development in typical and language delayed children

Children typically start to say their first words around their first birthday (Fenson et al., 1994; Fenson et al., 2007). At this age they already use a variety of gestures to communicate and they understand around 50 words (Fenson et al., 1994).

It is well reported that language comprehension precedes spontaneous language production (Fenson et al., 1994). By language production we are referring to the spontaneous use of language rather than imitation. Infants recognise their own name from 4.5 months (Mandel, Jusczyk, & Pisoni, 1995) and make associations with frequent words relating to people (e.g. mommy, daddy) or other common nouns as early as six months of age (Bergelson & Swingley, 2012; Tincoff & Juscyzk, 1999). At around 9 months, children show comprehension of words for games and routines (Syrnyk & Meints, 2017). At around 12 months, studies show that children are capable of mapping words to animate and inanimate objects but only match words (e.g. bird) with typical exemplars (e.g. sparrow) and are not yet able to extend the word to fit the entire category (Meints, Plunkett, & Harris, 1999; Southgate & Meints, 2000). From 18 months, children associate nouns and prepositions with atypical exemplars (Meints, Plunkett, Harris, & Dimmock, 2002; Meints, Plunkett, Harris, & Dimmock, 2004). Most children start to produce their first words at around 12 months (Fenson et al., 1994) and their vocabulary increases gradually until around 18 to 24 months when most children experience a vocabulary spurt (e.g., Reznick & Goldfield, 1992). Grammar also emerges at this age as most children start to produce verbs and two-word combinations (semantically underspecified, syntactic frame-slotting or agent-focus), in other words simple sentences which only contain content words (Bochner & Jones, 2003; Chan, Meints, Lieven, & Tomasello, 2010; Meints et al., 2004).

Children who do not achieve these milestones at the correct time are characterised as *language delayed* – language delay can be detected between 18 and 35 months of age (Rescorla, 2011). The most common delay is an *expressive language delay* (prevalence: 16% at 2 years), also referred to as late-talking, which affects the ability to produce language, at the same time that language comprehension (receptive language) is typical (Law, Boyle, Harris, Harkness, & Nye, 2000b). Some children also suffer from mixed delays (prevalence: 2.6%) in which both domains (receptive and expressive) are affected (Law, Boyle, Harris, Harkness, & Nye, 2000b). Such children have been found to have the poorest long-term outcomes (P. A. Silva, Williams, & McGee, 1987).

There is some research which indicates that early intervention for children with language delay at around 2 years of age can significantly reduce the risk of language problems in the preschool years (e.g., Buschmann et al., 2008; Buschmann et al, 2015). It has also been shown that direct language intervention by a speech and language therapist was not superior compared to parent-based interventions in terms of language progress over time, in fact children in the two groups did not differ significantly from each other at any testing point during the preschool years (Baxendale, 2003). Parent-based language intervention programs, such as the Heidelberg Parent-based Language Intervention (HPLI), consist of teaching parents language facilitation strategies which subsequently aim to help children improve their language skills (Roberts & Kaiser, 2015). Importantly, the HPLI has also shown to be cost and time effective (Buschmann et al., 2008). Furthermore, oral language skills could also be improved in UK nursery settings, in which teaching assistants conducted small group interventions over a period of 20 to 30 weeks (e.g., Fricke, 2013, Fricke et al. 2017). Six months later, improvements were still visible when comparing the intervention with the control group. As speech and language therapy is not commonly offered to children below 3 years of age (Whitehurst & Fischel, 1994), parent-based or nursery lead interventions seem particularly suitable for language delayed toddlers.

The reasons for language delay can be very different. For example, children may experience disruptions to language acquisition due to genetic (e.g. Williams Syndrome), social (e.g. isolation or personality disorder), sensory-perceptual (e.g. deafness) or neurocognitive (learning disability or brain damage) factors (Tartter, 1998). However, language problems can also occur in children who do not show any other than language-specific problems (Desmarais, Sylvestre, Meyer, Bairati, & Rouleau, 2008). Thus, there is a difference between children whose delays are secondary to other conditions, for example Autism Spectrum Disorder (ASD) or general developmental disabilities and those children with a primary language delay with no underlying medical condition (Law, Boyle, Harris, Harkness, & Nye, 2000b).

Several factors have been detected which put children at an elevated risk to develop a primary language delay such as preterm births (e.g., D'Odorico, Majorano, Fasolo, Salerni, & Suttora, 2011; Guarini et al., 2010; Stolt, Haataja, Lapinleimu, & Lehtonen, 2009), recurring ear infections (e.g., J. E. Roberts, Burchinal, & Zeisel, 2002; J. E. Roberts, Rosenfield, & Zeisel, 2004), a family history of dyslexia (Soriano-Ferrer & Piedra Martínez, 2017) or speech and language problems (e.g., Bishop, Price, Dale, & Plomin, 2003; Zambrana, Pons, Eadie, & Ystrom, 2014).

Furthermore, many late talkers (50-75%) catch up with their peers, whilst others develop language disorders and some children with no initial delay develop language disorders later (e.g., Bavin & Bretherton, 2013; Paul, 2000; Rescorla, Dahlsgaard, & Roberts, 2000; Roos & Ellis Weismer, 2008). However, even among recovered children many show low scores, albeit within the normal range, and underperform on language and reading tests compared to matched controls during the school period (e.g., Dale, McMillan, Hayiou-Thomas, & Plomin, 2014; Rescorla, 2002).

When school-age children with language disorders were studied, these studies generally showed delays from the outset of language acquisition (Pickles et al., 2009). It has been stated repeatedly that early intervention could result in fewer of the social, emotional and mental health issues that are often associated with language impairments during childhood and better academic outcomes through adolescence and adulthood (St Clair, Pickles, Durkin, & Conti-Ramsden, 2011; Whitehouse, Watt, Line, & Bishop, 2009). Hence, researchers have suggested that language intervention should be made available as early as possible (Buschmann et al., 2008). This is because early language intervention during the first two years of life has shown to significantly improve language skills (Buschmann et al., 2009; Buschmann, Multhauf, Hasselhorn, & Pietz, 2015; Ciccone, Hennessey, & Stokes, 2012; M. Y. Roberts & Kaiser, 2015) and predict language and cognitive outcomes at three years (Rodriguez et al., 2009). Identifying and treating children in need for additional support as early as possible may protect them from developing academic and behavioural problems as a result of impoverished communication skills.

There are no commonly cited language delay theories which emphasize the importance of prediction. This could be because the causes of language delay and its trajectory are not yet fully understood. Nevertheless, prediction – as a scientific method- has been considered important for making evaluations about theories in the physical sciences for centuries (Hofman, Sharma & Watts, 2017). Furthermore, prediction is also used in medical and psychological research in order to investigate psychometric properties of (e.g. screening) tests and make best possible clinical outcome assessments for patients (McClimans, Browne & Cano, 2015). This is particularly useful for my PhD study as I am not only interested in group-level differences but also to investigate future language status of individuals with initial language delay.

1.2. Risk factors in language development

Different factors influence language development and will be discussed in the following sections. Risk factors such as SES, prematurity and risk of dyslexia should also be taken into account when studying the predictive validity of the UK-CDI and will be further discussed when reviewing the literature of the predictive validity of other CDIs.

There is a high variability in the rate of language development when comparing children within and across languages. Differences in vocabulary development have been described within and across children of the same language community (e.g., Dale, Price, Bishop, & Plomin, 2003; Fenson et al., 2007). Cross-linguistic research has also established significant differences in the rate of vocabulary growth between languages (e.g., Bleses, Vach et al., 2008; Caselli, 1995; Hamilton et al., 2000). This study investigates monolingual children's development, thus the impact of additional language environments will not be discussed here.

The presentation of risk factors below is not intended to be fully comprehensive, but it covers the most commonly studied risk factors (family history of speech and language delay, male gender, parent education levels, preterm birth (perinatal factors)) as well as some less studied factors (ear infections (childhood illnesses), later birth order, SES) according to Reilly et al. (2007).

1.2.1. Ear infections

Ear infections which occur with sticky or thick fluid in the middle ear are referred to as otitis media with effusion, OME. Hearing is a requirement for typical language acquisition, however, this is often temporarily impaired in children with OME. This is problematic as those infections increasingly occur between 1 and 4 years of age, the critical period for language development (Maw, Wilks, Harvey, Peters, & Golding, 1999). It has been suggested that repeated OME outbreaks are associated with less advantageous speech and language outcomes (Maw et al., 1999; J. E. Roberts et al., 2002; Shriberg, Friel-Patti, Flipsen, & Brown, 2000). However, a meta-analysis of studies between 1966 and 2002 showed no or very small relationships between OME during early childhood and children's prospective speech and language outcomes during the preschool period (J. E. Roberts et al., 2004). There is little consensus amongst CDI researchers in terms of exclusion criteria for ear infections. For the creation of CDI norms, some researchers retained data of children with repeated ear infections (e.g., Eriksson & Berglund, 1999; Fenson et al., 1994; Fenson et al., 2007) and others excluded them (e.g., Kalashnikova, Schwarz, & Burnham, 2016; Kern, 2007). Other researchers included children with otitis media who received a tympanostomy tube to prevent the build-up of fluid in the middle ear (Bleses et al., 2008; Wehberg et al., 2007).

1.2.2. Family history of speech and language problems

Language delay and persistent language problems are heritable (e.g., Bishop et al., 2003; Zambrana et al., 2014). There is also a genetic influence on early reading abilities (Hohnen & Stevenson, 1999) and dyslexia (Soriano-Ferrer & Piedra Martínez, 2017) so that about 50% of children who have parents with dyslexia will also have dyslexia (van der Leij, Lyytinen, & Zwarts, 2001). Poor early language skills have also been shown to predict dyslexia (Gallagher, Frith, & Snowling, 2000; Scarborough, 1990). Traditionally, researchers argued that a phonological deficit is the cause for dyslexia (e.g., Stanovich, 1988) but more recent research (Pennington, 2006) suggests that a combination of causes (e.g. phonological deficit and problems with processing speed) lead to dyslexia. Whilst genetic factors play a strong role in developing dyslexia, research has also shown that environmental factors such as shared book reading can serve as protective factors by reducing the risk of poor reading skills (Torppa et al., 2007). Recent studies found that the stability of early language development for children with familial risk of language or reading problems was stronger compared to children without familial risk (Unhjem, Eklund, & Nergård-Nilssen, 2015; Zambrana et al., 2014). These at-risk children may be disadvantaged from the outset of language development in regards to slower development in linguistic and related domains (e.g. auditory processing skills, speech perception, phonological awareness) (e.g., Boets et al., 2011; Schaadt, Mannel, Meer, Pannekamp, & Friederici, 2016).

Lessened language abilities are found in children with familial risk of dyslexia already from two months in speech perception tasks (van Zuijen, Plakas, Maassen, Maurits, & Leij, 2013). They are at greater risk for language delay (P. Lyytinen, Poikkeus, Laakso, Eklund, & Lyytinen, 2001; P. Lyytinen, Eklund, & Lyytinen, 2005) and for developmental language disorder (DLD) (i.e. developmental language disorder formerly referred to as specific language impairment, SLI) at 4.5 years (Nash, Hulme, Gooch, & Snowling, 2013).

When investigating the contributions of mothers and fathers separately, paternal but not maternal experiences of histories of speech, mental retardation and learning difficulties were associated with DLD (Tomblin, Smith, & Zhang, 1997).

Furthermore, the co-occurrence of language / reading development problems with motor problems (e.g., Bishop & Edmundson, 1987; Kaplan, Wilson, Dewey, & Crawford, 1998; King-Dowling, Missiuna, Rodriguez, Greenway, & Cairney, 2015) but also with other disorders such as ADD/ADHD (e.g., Kaplan et al., 1998) has been reported. Children at risk for dyslexia and slow motor skills had significantly lower vocabulary and grammar scores than controls or at-risk children with better motor skills when tested during their second year of life. However, control children without risk of dyslexia and with slow motor development were not at risk for language delay (Viholainen, Ahonen, Cantell, Lyytinen, & Lyytinen, 2002). A common genetic origin has been proposed as an explanation for the co-occurrence of dyslexia and motor problems (Gilger & Kaplan, 2001).

1.2.3. Gender

As has been shown in numerous norm-based studies across different languages, girls acquire language at a faster rate than boys but the effect sizes are usually small (e.g., Bleses et al., 2008; Fenson et al., 1994; Fenson et al., 2007; Simonsen, Kristoffersen, Bleses, Wehberg, & Jørgensen, 2014; Szagun, Steinbrink, Franik, & Stumper, 2006). In a recent re-analysis of CDI Infant and Toddler data of 10 language communities, it was confirmed that the gender differences were robust, as no interactions between gender and language community were detected (Eriksson et al., 2012).

US-American CDI data suggests that gender differences exist for all CDI subscales but the magnitude of the difference was very small accounting for around 1-2% of the variance in the Infant version and around 1-3% for the Toddler version (Fenson et al., 1994; Fenson et al., 2007).

Furthermore, even though CDI studies found gender differences, no differences were detected in the Swedish Infant and Toddler CDI norms for any subscales (Berglund & Eriksson, 2000; Eriksson & Berglund, 1999).

Whilst in most CDI studies, total vocabulary, gesture and grammar scores are compared between boys and girls, it has also been found that the composition of early lexicons differ significantly between girls and boys (Wehberg et al., 2008). Whilst girls start to use words earlier overall, they use words earlier in particular for words related to social relations and personality, and objects to be cared for. Boys, in contrast, name words earlier for "loud moving objects, objects they can act on and certain food-related items" (Wehberg et al., 2008, p. 81).

The reason for gender-specific differences may be biological in terms of neuropsychological research indicating earlier brain maturation in girls compared to boys (Lenroot et al., 2007), or environmental reasons may be responsible for the early advantage for girls. For example, it has been shown that children show preferences for gender-typed toys during infancy (Alexander, Wilcox, & Woods, 2009; Campbell, Shirley, Heywood, & Crook, 2000; Todd, Thommessen, & Barry, 2016) and that female-typed toys encourage more language interaction (Caldera, Huston, & O'Brien, 1989). In addition, play behaviour by caregivers has shown to be more symbolic rather than exploratory with girls (Suizzo & Bornstein, 2006) - this is important as symbolic play is strongly correlated with language development (e.g., Orr & Geva, 2015). This early advantage seems to linger as girls outperform boys on early literacy measures up to 3rd grade (Lee, 2011). Furthermore, Bleses et al. (2016) found that correlations between the expressive vocabulary subscale of the Danish CDI: Words and Sentences collected between 16 – 30 months with language and literacy tests at sixth grade (around 12 years) were significant and stronger for boys compared to girls. In contrast, Hohm et al. (2007) investigated expressive and receptive language skills at 10 months -in a sample of high variability in risk for future language problems - and found that these abilities significantly correlated with educational outcome and cognitive scores 10 years later. They also reported that prediction was better for girls compared to boys, particularly for verbal measures.

The reason for this difference in results is yet unclear and may be due to the age of first measurement and gender as suggested by Bleses et al. (2016). Overall, gender differences are likely to exist but effects are small which explains that some studies did not find significant differences between boys and girls (Rescorla & Dale, 2013).

1.2.4. SES

Associations between socio-economic status (SES) and language ability have been demonstrated, usually with better results in children from higher socio-economic groups (e.g., Arriaga, Fenson, Cronan, & Pethick, 1998; Hoff-Ginsberg, 1998; Tomblin, 2012). The reason for this could be that SES may indirectly affect language outcomes through the impact of family characteristics. For example, it was found that maternal language (i.e. vocabulary, mean length of utterance, conversational skills) mediated the relationship between maternal education/occupation and child language (Bornstein, Haynes, & Painter, 1998; Hoff, 2003). Apart from parent-child interactions, other factors such as child characteristics are also important. For example, lower language processing skills were found in 18-month-olds from low SES families compared to high SES children as suggested by Fernald et al. (2013). Their research found a 6 months gap at 24 months of age in terms of vocabulary and language processing efficency between high and low SES children.

In addition, the environment also influences language development. For example, it has been shown that children from low SES families had fewer resources available at home (e.g. number of books, stimulating toys see Bradley et al, 2001; Rodriguez et al., 2009; Froiland et al., 2013) and were less likely to access a variation of resources in the community (e.g. zoos, modern libraries or parks see Pogash 2016; Neuman & Celano, 2011) which help to develop broader vocabularies and language growth and put children from high SES families at an advantage.

It is also well established that SES correlates with children's performance on language tests during primary school (e.g., Lee, 2011; NICHD Early Child Care Research Network, 2005; Norbury et al., 2016) and secondary school (e.g., Spencer, Clegg, & Stackhouse, 2012) and prevalence of language impairments is higher in low SES backgrounds (Tomblin, Smith et al., 1997).

Different researchers operationalised SES using different measures such as household income (e.g., Arriaga et al., 1998; Lee, 2011), maternal education (Fenson et al., 1994; Fenson et al., 2007), maternal occupation (Berglund, Eriksson, & Westerlund, 2005), or postcode analysis using the index of multiple deprivation (e.g., Duff, Nation, Plunkett, & Bishop, 2015; Norbury et al., 2016). The reason for choosing data other than household income in order to measure SES lies in the difficulty of gaining reliable income information (Vernon-Feagans et al., 2008). Furthermore, comparing household income information from single and two parent families is difficult and may be less informative than using the more stable status of the primary caregiver's education. However, both maternal education and household income are commonly used in language development research. CDI studies often use maternal education as a proxy for SES as first suggested by the original norming study (Fenson et al., 1994). Long-term predictive validity studies using CDIs found that low maternal education and male gender were good predictors for future language outcomes (e.g., Ghassabian et al., 2014; Reese & Read, 2000; Sachse, Saracino, & von Suchodoletz, 2007). The UKBTAT study (UK Bilingual Toddler Assessment Tool, see Floccia, Sambrook, & Delle Luche, 2017) employs the same UK-CDI Family Questionnaire as used in the UK-CDI norming study (Alcock et al., 2017; Alcock et al., in prep) and in the current study and investigates the vocabulary development in English children growing up with an additional language in the UK. In their study, family household income (amongst other indicators of SES, for example maternal education, number of bedrooms, language exposure and other demographic, childcare and medical information) was the best predictor for language ability in 2-year-old bilingual children (Floccia et al., 2017).

When using the American CDI (Fenson et al., 1994), low-income children between 16 and 30 months were at a disadvantage for all vocabulary and grammar measures in comparison to middle-income families (Arriaga et al., 1998). Furthermore, American English research studying children in naturalistic interactions at two years found that productive vocabularies grew at faster rates in high-SES compared to mid-SES children over a 10-week period (Hoff, 2003). However, with the Portuguese Infant CDI (Short Form), SES differences were only detected in early word comprehension, favouring children from mid-SES families over low as well as high-SES families (Frota et al., 2016). Other studies detected no association between SES and language, for example, using the Swedish CDI at 18 months (Berglund et al., 2005). On the other hand, other studies reported higher scores for children with low SES on the NZ Toddler CDI at 19 months for production (Reese & Read, 2000), the Australian Infant CDI at 12 months for comprehension (Bavin et al., 2008) and the American Infant CDI at around 12 months for comprehension and production (Feldman et al., 2000), so a mixed picture arises. If no SES differences were detected, this may be due to less diverse sampling as many studies did not include representative numbers of parents with very low or no formal qualifications.

The diverging results could be the lack of representative sampling as suggested by Bleses et al. (2016) as they reported different results depending on the samples used for the Danish Toddler CDI. It also possible that cultural or ethnic differences are important here. For example, African American parents from predominantly low-income backgrounds were more likely to underreport vocabulary and grammar on the US CDI (Fenson et al., 1994) when compared to other standardised instruments (Sequenced Inventory of Communication Development - Revised, see Hedrick, Prather, & Tobin, 1984b) in a longitudinal study following children between 18 through 30 months (J. E. Roberts, Burchinal, & Durham, 1999). The expectations of achieving different milestones may differ depending on maternal education levels (Feldman et al., 2000) and they may also vary depending on cultural or ethnic groups. The structure of countries' demographics may also play a prominent role, as for example, the gap between socio-demographic groups is very narrow in Sweden which could explain the results by Berglund et al. (2005).

1.2.5. Birth order

As the child's environment influences the way children develop language (e.g., Collisson et al., 2016; Korpilahti, Kaljonen, & Jansson-Verkasalo, 2016), it is not surprising that birth order plays a role. First-borns receive more child-directed speech from their primary caregivers, which also differs in quality in speech directed towards them compared to later-born children, in other words mothers of first-borns use longer utterances and fewer questions (Hoff-Ginsberg, 1998). In contrast, later-borns benefit from the opportunity to overhear speech between the caregiver and older toddlers which is not as complex as conversations between two adults (Hoff-Ginsberg, 1998).

This leads to different acquisition patterns in first-borns compared to later-borns. Early language differs quantitatively - first-borns understand more words at 18 months (Berglund et al., 2005) and acquire their first 50 words earlier than their later-born siblings (Pine, 1995). Pine's study (1995) also found very strong correlations between sibling pairs and early vocabulary suggesting a strong genetic connection. Other studies found that the biological factor of gender or gestational age plays a more important role than the social factor of birth order in language development (Berglund et al., 2005; Kern & Gayraud, 2007).

Birth order also impacts on the composition of the first 100 word vocabularies. A Danish CDI study demonstrated that first-borns used more terms for sound effects and produced names for mother, father, babysitter and work earlier. In contrast, later-borns used words earlier that normally occur in older children's dictionaries (e.g. scissors, dance, read, afraid), words for brother or sister and actions with another person (e.g. bite, hit, get) (Wehberg et al., 2008).

In terms of grammatical development, Hoff-Ginsberg (1998) found that first-borns were more advanced in terms of mean length of utterance (MLU) between 18 and 29 months in a study of natural interactions during different daily situations between child and caregiver. The study also investigated conversational skills and found that first-borns used more expansions (i.e. child adds new information to mother's prior utterance or introduces a new, related theme), more non-contingent responses (i.e. child does not maintain the topic of mother's previous utterance) and fewer routines than their later-born counterparts. Later-borns used significantly more *frozen phrases* (e.g. that's mine) during their first 100 word phase (Pine, 1995) and words for first and second-person pronouns between 21 and 24 months of age (Oshima-Takane, Goodz, & Derevensky, 1996). A likely explanation for later-born children's ability to correctly use personal pronouns earlier may lie in their opportunities to overhear the use of pronouns in more varied contexts in triadic interactions in which an older sibling is present. This may be more favourable than dyadic interactions as caregivers have shown to frequently replace pronouns by proper names in child-directed speech in order to simplify language for younger children (Oshima-Takane et al., 1996).

In summary, whilst it might be true that parents of later-born children complete questionnaires differently due to time-constraints or less attention to the development of their later-born child (Bornstein, Leach, & Haynes, 2004), these findings also show socialization differences and reflect differential input towards children depending on their birth order status (Hoff-Ginsberg, 1998).

1.2.6. Preterm birth

Worldwide estimates of preterm births are around 11%, but the prevalence is only around 5% in some northern European countries (Blencowe et al., 2012). Full term birth is classed as ranging between 37 and 42 weeks and birth before 37 weeks of gestation is classified as premature (Blencowe et al., 2012). The highest rates of preterm births (i.e. 84%) occur in moderate-to-late preterms (born at 32 – < 37 weeks of gestation) (Blencowe et al., 2012).

Poorer language skills have been reported in preterm compared to full-term children in terms of preverbal communication (e.g., Crnic, Ragozin, Greenberg, Robinson, & Basham, 1983), early receptive vocabulary (e.g., Stolt et al., 2009), phonology (e.g., D'Odorico et al., 2011; Sansavini et al., 2007), grammar and literacy (e.g., Guarini et al., 2010). Monson et al. (2018) investigated brain differences in preterm and full-term children to find precursors of language problems in preterm children. They found that during the perinatal stage children who were born preterm had less developed auditory systems and that full-term children benefited from major auditory developments during the last weeks of pregnancy (between 26 to 40 weeks). Furthermore, the research team found a link between a less mature nonprimary auditory cortex and language delay at 2 years in preterm children.

CDI studies investigating differences between preterm and full-term children have focused on premature children with very low birth weight (Stolt et al., 2007; Stolt et al., 2009) and on very premature children (born at 28 – < 32 weeks of gestation) (Sansavini et al., 2011; Suttora & Salerni, 2012). Other research also investigated the differences between full-term, moderate-to-late preterm, very preterm and extremely preterm (<28 weeks) children in terms of language development (e.g., Kern & Gayraud, 2007; Putnick, Bornstein, Eryigit-Madzwamuse, & Wolke, 2017). Kern and Gayraud (2007) using the French CDI: Words and Sentences found that children born moderate-to-late preterm did not differ from full-term children at 24 months in vocabulary size, grammatical categories and maximum length of utterance, in contrast to extremely preterm children who performed significantly poorer on all lexical and grammatical measures.

Putnick et al. (2017) investigated long-term stability between 5 months and 8 years of language skills between very preterm, moderate-to-late preterm and full-term children. They found that very preterm children's language scores were the lowest of all groups at all testing points, and with the most stable language between 20 months and 8 years even after controlling for cognitive skills and family SES. Furthermore, language skills between moderate-to-late preterm and full-term children did not differ at 5 months and 8 years, but at 20 months, 4 years and 6 years – here, moderate-to-late preterm showed significantly lower language scores than full-term children.

A recent study found that 36% of moderate-to-late preterm born infants scored below 1 SD from the mean on the language scale of the Bayley-III in comparison to 16% of the typical population at 24 months (Spittle et al., 2017). Of these moderate-to-late preterm children, 4% fell below 2 SD in comparison to around 2% of the typical population. Even though children born moderate-to-late preterm showed more language delay, they were more likely to experience moderate delays rather than severe delays. However, motor development delays were most common amongst children born moderate-to-late preterm as 42% scored below 1 SD and 8% scored below 2 SD from the mean in comparison to 16% and 2% for typical children, respectively. Interestingly, the prevalence of children with cognitive delays was similar using 1 SD as cut off for moderate-to-late preterm (19%) and typical children (16%); however, more moderate-to-late preterm children (6%) showed severe cognitive delays (below 2 SD from the mean) than typically developing children (2%).

Exclusion criteria for premature children were more stringent in the original CDI version compared to subsequent CDIs in other languages. This may be due to diverging results in the literature or improved obstetric and neonatal care. For example, children were excluded who were 6 or more weeks (≤ 34 weeks) premature for the US MB-CDI (Fenson et al., 1994; Fenson et al., 2007) whilst for the Danish CDI children born below 32 weeks of gestation were excluded (e.g., Bleses et al., 2008a; Wehberg et al., 2007).

1.2.7. Summary

Overall, environmental and biological factors are correlated with language development, but these factors could only explain small amounts of variance in late talking status, whilst biological factors were generally somewhat better predictors than environmental factors (Rescorla & Dale, 2013). However, in a twin study it was found late-talking has a higher heritability at 3 years than at 2 years when the influence of the shared environment was stronger than genetic endowment (Dale & Hayiou-Thomas, 2013). It can be concluded that there is still a lot of variance in later language status which cannot be explained by the commonly used risk factors.

In the next chapter we will focus on the central questions of this PhD project - questions of measurement and prediction of language abilities in children. How can we measure early language development and differences in language abilities? Can we make predictions from these measures?

Hence, the literature on measuring early word knowledge focusing on Communicative Development Inventories will be presented and summarised. I will then present studies that have investigated the predictive validity of CDIs in other languages than British English or in non-normed British CDIs.

2. Communicative Development Inventories (CDI)

Communicative Development Inventories (CDIs) are parent-report instruments about children's early communication and language. They capture the typical course and the naturally occurring variability in language development (Fenson et al., 1994; Fenson et al., 2007). All CDIs consist of a vocabulary checklist (original US MacArthur-Bates CDI:WG (Infant form): 396 items, US CDI:WS (Toddler form): 680 items, US CDI-III: 100 items), additional understanding of phrases and gesture scales for the infant form (US CDI:WG used between 8 - 18 months) and grammar scales for the toddler form (US CDI:WS used between 16 - 30 months) and the US CDI-III (30 – 42 months) (e.g., Dionne, Dale, Boivin, & Plomin, 2003; Fenson et al., 2007; Law & Roy, 2008). Due to their size, for some languages, short forms have been created for the infant and toddler versions (e.g., Fenson et al., 2007).

CDIs have been adapted from previously existing word lists such as the Early Language Inventory, ELI, (Bates, Bretherton, & Snyder, 1988) and the Language and Gesture Inventory (Bates et al., 1986). Many studies investigating the predictability of infants' and toddlers' language have made use of a range of parent-report questionnaires to assess early language development (e.g., for Early Language Inventory (ELI) see Bornstein & Haynes, 1998; for Communicative Development Inventory (CDI) see Feldman et al., 2005; for Language Development survey (LDS) see Klee et al., 1998; for Language Use Inventory (LUI) see Pesco & O'Neill, 2012; for Communication and Symbolic Behaviour Scales (CSBS) see Wetherby, Goldstein, Cleary, Allen, & Kublin, 2003).

Of these tools, Communicative Development Inventories (CDI) are the most commonly used parent-report instruments. Since their first publication in 1994 (Fenson et al., 1994), they have been adapted into and used in more than 60 languages (Fenson et al., 2015). Such parent-report questionnaires have been shown to have good psychometric properties and are quick to administer and cost-effective (Fenson et al., 1994; Fenson et al., 2007). Parents are usually able to fill in the questionnaire on their own and are a valid and reliable source of information in terms of their child's ability (Fenson et al., 1994; Fenson et al., 2007). CDIs may also be administered by the experimenter in interview format, for example, for illiterate parents (Alcock et al., 2015).

Furthermore, the experimenter does not need to interfere and elicit any language directly from the child as is typical during standardised language tests. In such situations, elicitation of language production may be difficult and yield different results depending on the temperament of the individual child, their willingness to cooperate in the given situation and the ability of the researcher to build a rapport with the child. In addition, experimental studies (e.g. EEG, ERP, eye-tracking) or observational studies can be time-consuming and costly compared with parents filling in language questionnaires.

2.1. Purpose of CDIs

Fenson et al. (2007) propose that CDIs can be utilized for a range of different research purposes. They can be used as a baseline measure, or to help select appropriate research stimuli, for example for semantics and grammar research (Meints, Plunkett, Harris, & Dimmock, 2004; Chan, Meints, Lieven, & Tomasello, 2010), to preselect children at different stages of language development (e.g., Gershkoff-Stowe & Smith, 1997), or to select those with specific language characteristics (see Thal, Bates, Goodman, & Jahn-Samilo, 1997 for late or early talkers) or atypical language profiles (see Thal, Bates, Zappia, & Oroz, 1996 for children with low word combinations despite high vocabulary size). CDIs are also useful to match participants on language skills (e.g., McGregor, Sheng, & Smith, 2005) or for investigating other factors associated with language development (e.g., Collisson et al., 2016).

Their ease of use for large-scale projects has also enabled the measuring of the impact of key environmental, societal and biological factors on language development (e.g., Dionne et al., 2003). CDIs were also important in highlighting that language is not merely a maturational (i.e. age-related), but also a developmental process, for example in 2-year-old children, while grammar develops as children mature, grammar is more strongly related to lexicon size than to age (McGregor et al., 2005).

CDIs have been used for different populations and purposes which helps to further our understanding of the early development of language. For example, other studies have investigated the trajectory of different and atypical populations, such as bilingual children (e.g., Cattani et al., 2014; Rinker, Budde-Spengler, & Sachse, 2017) or used CDIs for cross-linguistic research (e.g., Bleses, Vach et al., 2008b; Wehberg et al., 2007).

In addition, the authors of the original US MacArthur-Bates CDI propose the use of parent-report checklists for screening for language delay (e.g., Klee, Pearce, & Carson, 2000) as has been implemented for different languages and age groups (e.g., Bleses, Vach, Jørgensen, & Worm, 2010; Sachse, Saracino et al., 2007). CDIs can also be used to evaluate older children with language skills in the range of the CDI for children with language impairment (e.g., Thal, O'Hanlon, Clemmons, & Fralin, 1999) or other special populations (e.g., Yoder, Warren, & McCathren, 1998). Furthermore, Fenson et al. (2007) suggest to use the CDI to detect "aspects of the child's communicative skills that may be targeted for intervention" (p. 43) (see also Ciccone et al., 2012) or to evaluate the efficiency of interventions (see also Buschmann et al., 2009). CDIs can also help to monitor the language progress of children with language delay (Thal et al., 1999).

In practice, CDIs are indeed used for screening for language delay as suggested by the authors (Fenson et al., 1994). Scores are compared against national averages in order to evaluate if a child falls within the typical range of the population norms. This is possible as standardised CDIs consist of norms for age bands in months. Health professionals or early education providers in different countries (e.g. Germany, Denmark) use age-appropriate CDIs in order to describe children's course of development between one and two years of age (see FRAKIS - Szagun, Stumper, & Schramm, 2009) or detect children who are likely to develop language delay between one and two years (Grimm & Doil, 2006), between two and three years (von Suchodoletz & Sachse, 2008; von

Suchodoletz, Kademann, & Tippelt, 2009) and at around 3 years of age (Vach, Bleses, & Jørgensen, 2010).

CDIs may be particularly useful for clinical purposes. Other widely used standardised in-person language tests are not appropriate due to lack of norms for children during early infancy (e.g. Reynell Developmental Language Scales III (Edwards, Fletcher, Gurman, Hughes, & Letts, 1997) were normed from 18 months to 7 years; the New Reynell Developmental Language Scales (Edwards, Letts, & Sinka, 2011) are only normed from 3 years onwards) or lack of British English norms (e.g. Sequenced Inventory of Communication Development- Revised: Hedrick, Prather, & Tobin, 1984a). An exception are the Preschool Language Scales-5UK (Zimmerman, Steiner, & Pond, 2011) for which norm data exist from birth to 7 years and 11 months for language in terms of auditory comprehension and expressive communication. Nevertheless, Fenson et al. (2007) emphasize the importance of using a range of language assessments and not just parent-report in order to provide a genuine picture of a child's language ability. Hence, the combined use of the UK-CDI in conjunction with other instruments such as the Preschool Language Scales-5UK may yield a comprehensive description of a child's communicative scope in clinical settings.

2.2. Critical assessment of CDIs

While CDIs are clearly useful as described above, there are also some pitfalls. There is large variability in children's language trajectories, and some researchers have cautioned about the potentially low predictive value of CDIs to detect language delay or disorders in children between 3 and 4 years of age (Dale et al., 2003; Feldman et al., 2005). The predictive validity of different CDIs will be discussed as part of chapter three. The following chapters focuse on general criticisms towards CDIs which have been stated since the introduction of CDIs (Fenson et al., 1994). We aimed to address these issues when constructing the UK-CDI:WG which will be discussed in the second part of this chapter.

2.2.1. Socio-economic status (SES) and ethnicity

Socio-economic status (SES) and ethnicity are important factors to consider. The original MacArthur-Bates CDI (MB-CDI) norm sample (Fenson et al., 1994) was biased towards middle class families. Hence, the norms were updated in 2007 to include mothers from more varied educational backgrounds (Fenson et al., 2007). However, the lowest education level (i.e. high school or less) is still under-represented and the highest educational level (i.e. college diploma and similar) is over-represented. Furthermore, the data show an over-representation of Caucasian families according to the 2000 US census. As research has shown that children with parents from lower SES families can have lower linguistic abilities compared to children from higher SES backgrounds (Hart & Risely, 1995; Ryan, Gibbon, & O'shea, 2016; Sharkins, Leger, & Ernest, 2017), one would expect this to be reflected in CDI scores. However, parents from lower socio-economic backgrounds and African American backgrounds compared to European Americans over-reported CDI: Words and Gestures scores for their children's very early word and sentence comprehension and in some cases word production (Feldman et al., 2000; Fenson et al., 2015; Reese & Read, 2000).

The CDI: Words and Gestures comprehension scale has been criticized due to such inflated scores during the early stages of language development (Tomasello & Mervis, 1994) and the difficulty of assessing word comprehension in contrast to word production and use of gestures (Stiles, 1994). The MB-CDI authors were also aware of this and advised caution when interpreting MB-CDI scores of children below the age of 1 year from low SES backgrounds (Fenson et al., 1994; Fenson et al., 2000; Fenson et al., 2007). Nevertheless, other studies found that parents are able to give accurate accounts of their child's early word comprehension when comparing CDIs with experimental studies at the laboratory (e.g., Mills, Coffey-Corina, & Neville, 1997; Styles & Plunkett, 2009; Syrnyk & Meints, 2017). This discrepancy between studies is likely due to sampling as most laboratory studies include Caucasian children from middle class backgrounds, compared to the studies which specifically aimed to test representative numbers in terms of SES and ethnicity.

Differences in terms of SES and ethnicity were also found on the MB-CDI: Words and Sentences. Over-estimations by parents with low SES diminished over the course of the second year (Feldman et al., 2000). At approximately 2 years, children with lower SES were reported to have lower word production scores and word combinations than children with high SES. Pan et al. (2004) found that CDI scores were not significantly associated with maternal education at 2 years, but instead with ethnic background – here, mothers who classed themselves as from a *White* ethnic background reported higher vocabulary scores compared to Black or Hispanic mothers, see also Kreisman (2012). Roberts et al. (1999) showed that at 30 months a very high proportion of children (45%) from primarily low-income African American families were reported to have scores below the 10^{th} percentile on the CDI: Words and Sentences (Short Form). The researchers showed here that parents under-reported their child's ability as became evident when compared to standardised in-person assessments.

This shows that clinicians and researchers need to be cautious when interpreting CDI scores from groups who are traditionally less studied in child language research. Reporting styles and vocabulary and grammar (e.g. optional use of some inflectional morphemes in Black English Vernacular (BEV)) may vary across subgroups and may differ depending on the age of the child (Feldman et al., 2000).

2.2.2. Instructions/ training

The original MB-CDI contains brief instructions at the beginning of each subsection (Fenson et al., 1994). For the first subsection First Signs of Understanding examples of behaviours are given which children may show when they understand frequently used short phrases by the caregiver (e.g. "no, no"). However, for the Vocabulary Checklist there are no detailed explanations given to parents about word understanding ("understands") and production ("understands + says"). The authors do not explicitly inform the respondent that "understands + says" means that children have to use the words in their own spontaneous speech and that imitations do not count. However, they do explain that parents should give the child credit for mispronounced words. In the User's Guide of the updated norms, the authors admit that different researchers distribute different instructions alongside the CDI which in their eyes improves the generalizability of the norms (Fenson et al., 2007). However, at the same time they compiled a long list with general instructions that parents may be provided with in addition to the CDI (e.g. "It is helpful to highlight the Vocabulary Checklist and to elaborate on the difference between the replies *Understands* and *Understands and Says*" (Fenson et al., 2007, p.16)). It is unclear why these points have not been integrated in an updated version of the MB-CDI.

As mentioned above, it is possible that parents are potentially more accurate when reporting gestures or produced words in contrast to word comprehension. This may be due to different decision-making strategies for the specific CDI categories. For gesture or word production, the respondent has to recall a situation in which the child used a gesture or said a word, whereas for word comprehension it may not be as clear as the parent has to understand and appreciate from their experience with the child which words their child understands. Therefore, it may be that the results are less consistent for word comprehension in the early stages of language development (see also Feldman et al., 2000; Stiles, 1994).

Some parents may accept a word as understood if the child seems to understand a word in highly constrained contexts (see Feldman et al., 2000; Tomasello & Mervis, 1994 for further information). For example, in a common eating situation when familiar foods (e.g. milk, banana) are available and one unfamiliar item (e.g. pasta) and the parent names the new word in an utterance (e.g. "let's eat spaghetti"). It is also possible that parents have difficulty distinguishing between word comprehension that is based on pure verbal input compared to the involvement of other non-linguistic or gestural cues, for example the parent asking the child to give them an object whilst the parent is looking at the object and holding their hands out in expectation (Chapman, 1978).

Another criticism of the UK-CDI as well as other CDI versions is that they do not explicitly define word comprehension in the instructions. It is possible that further instructions during an interview may be advantageous over the postal procedure used by the researchers in the norming study (Tomasello & Mervis, 1994). This is because the definition of comprehension was briefly explained at the beginning of the interview. For the UK-CDI parents were given the opportunity to contact the research team (via email, phone or letter), if they had any questions in regards to the completion of the UK-CDI. Furthermore, it was proposed that training, practice or feedback sessions for parents may be helpful in yielding more valid responses (Feldman et al., 2000).

2.2.3. Parents as informants

This leads to the next question of how well parents are able to remember their child's language (Stiles, 1994). Parents' reports about word comprehension were more stable for some categories than for others over a 2-week period on the CDI: Words and Gestures for children with developmental delay. Yoder et al. (1997) found that parents consistently reported comprehension of words for nouns, games and routines and action words on the word level, whilst the item-by-item stability was insufficient for descriptive words, prepositions, quantifiers and temporal terms. The authors conclude that children have more opportunities to respond to words like nouns in interaction with the caregiver in contrast to words like prepositions which usually co-occur with gestures by the speaker and it is unclear if the child understands the word or only the gesture. This means parents may find situations more memorable in which the child correctly executes an action requesting objects or actions (e.g. 'Give me a **biscuit'**, '**kick** the ball') than words that are usually short and uttered in complete sentences and make little sense if used in isolation (quantifiers, prepositions etc.). Other authors have also suggested that parents may find it difficult to report exact words, but they seem to give accurate accounts about their children's vocabulary size (Reznick & Goldfield, 1994; Styles & Plunkett, 2009).

Furthermore, it has been questioned if parents interpret the questions posed in the CDI in the same way and if they embellish their child's abilities (Stiles, 1994). If the reason why parents embellish their children's skills is because they want to answer questions in a socially desirable fashion, we would expect that parents with a stronger tendency to answer in a socially desirable way would tick more words on the CDI regardless of their child's actual ability; however, this was not confirmed in a study by Bornstein & Haynes (1998) who did not find a significant relationship between mothers' tendency to answer in a socially desirable fashion and their report. Together with other assessments of the concurrent validity, this may remove some scepticism towards the usefulness of the CDI; however as introduced above differences exist in the way some groups fill in the CDI (Feldman et al., 2000; Fenson et al., 1994).

Even though mostly mothers complete the CDI, this is not a prerequisite. The reporting style of different types of respondents (i.e. fathers, mothers, another familiar person to the child) was examined and it was found that even though mothers reported a significantly higher increase in word comprehension between 12 and 20 months than fathers or the other person familiar to the child, the comprehension and production scores between the three groups were strongly correlated at 20 months (Bornstein, Putnick, & De Houwer, 2006). This means that children whose fathers completed the CDI may be at a disadvantage as their norm-referenced scores will be lower than they potentially should be. This may be because CDI norms are overwhelmingly based on mothers as reporters. It is unclear if a reporter bias is in this case due to social or biological differences in mothers and fathers or due to differences in hours of childcare (i.e. mothers usually spend more time with their children than fathers, especially when children are young) lead to differing amounts of opportunities in observing and judging language developments.

2.2.4. Construction properties of the CDI

According to the norming study, the pretend objects scale in the gesture subsection has caveats (Fenson et al., 1993, p. 52) as reported by the authors and was therefore removed in the updated norms (Fenson et al., 2007) and other subsequent CDI adaptations such as the UK-CDI questionnaire.

Furthermore, Feldman et al. (2000) criticised the original US-CDI (Fenson et al., 1994) as the authors only reported percentiles and no standard scores in the manual. Percentile ranks show individuals' positions in regards to the norm sample; however, they cannot show the amount of difference between scores (Anastasi & Urbino, 1997). This is not so severe for scores around the median, but becomes apparent in the difference of raw scores at the extreme ends of the distribution (Bailey & Wolery, 1989); for example the comprehension scores for girls at 12 months differed in 7 words between the 5th and 10th percentile, but is 103 words between the 90th and 95th percentile (Feldman et al., 2000). In order to be able to compare groups or test the effectiveness of an intervention, Feldman et al. (2000) created standard scores (i.e. z-scores, standard deviations and means) for the different subscales of the CDI: Words and Gestures and CDI: Words and Sentences from the sample in their large-scale study. In response to this, Fenson et al. (2000) explain that z-scores are not suitable as they underlie the assumption for linearity and normality. Particularly for the CDI: Words and Gestures, the data is naturally skewed with few words at the beginning and a rapid growth during the later stages of development. However, Fenson et al. (2000) agree that percentile ranks can be misleading in the early stages of development and advise against their use for age groups for which skills are just emerging. It should be noted that the updated norms include data on the descriptive statistics (mean, median, standard deviation, standard error of the mean, maximum and minimum score). Nevertheless Fenson et al. (2007) emphasize that percentiles best represent the data.

In addition, several authors support the construction of gender specific norms to make better distinctions between typical and slow developers within one gender group (e.g., Bleses et al., 2010; Fenson et al., 1994; Fenson et al., 2007). As pointed out above other factors (e.g. prematurity, family risk, birth order, SES, ear infections) are also important in interpreting CDI measures but norms are usually not created for these subgroups.

2.3. Introduction of the UK-CDI: Words and Gestures

Since the creation of the first CDIs by Fenson et al. (1994), these checklists have been adapted into many other languages and adaptations have been created for different varieties of English, for example a CDI:WS for New Zealand English (Reese & Read, 2000) and CDI:WG and CDI:WS versions for British English (Alcock et al., in prep; Hamilton et al., 2000; Harris, Law, & Roy, 2005; Klee & Harrison, 2001). Over the years, several research sites across the UK created subsequent adaptations from the US CDI:WG (e.g., Hamilton et al., 2000). However, until recently there was no standardised data available for UK children between 8 and 18 months. This gap has now been filled by the standardisation of the UK-CDI: Words and Gestures with population norms for children between 8 and 18 months (Alcock et al., 2017; Alcock et al., in prep), but standardised UK data and norms for toddlers between 16 and 30 months (CDI: Words and Sentences) assessing vocabulary and early grammar development are still outstanding.

The standardised UK-CDI infant version was created after permission had been granted from the MacArthur-Bates CDI Board. The newly developed UK-CDI: Words and Gestures (Alcock et al., in prep) has been adapted from the original US MB-CDI (Fenson et al., 1994; Fenson et al., 2007). It has been standardised and normed for UK children aged between 8 and 18 months. Like the original infant version, the UK-CDI comprises of three subsections: 1) understanding of first words and phrases, 2) vocabulary checklist for comprehension and production and 3) gesture scale. The psychometric properties of the UK-CDI have been assessed in terms of internal consistency, test-retest reliability, online- and paper version, and concurrent validity using in-person tests, namely the Preschool Language (Zimmerman, Steiner, & Pond, 2014) and an Object Selection Task. CDI norms allow the examination and comparison of key factors on language development with other variables, for example gender, ethnic background or socio-economic variables.

Amongst the aims of the UK-CDI project was to recruit a representative sample in terms of SES. In response to previous criticism that instructions of the CDI were not clear enough, focus groups comprising of low SES families were consulted during the pilot stage of the project. These served to gain feedback and improve instructions as well as render the appearance of the questionnaire more appealing to respondents. The objective was to collect sufficient answers also from low SES parents and implement their feedback, as improving instructions can help parents to answer as accurately as possible.

In order to gain more accurate results, each category of the CDI included an area for feedback on different or local words with the same meaning. Furthermore, regional variations were also offered for some words (e.g. baby/bairn/wee one). The UK-CDI also enables researchers to study the developmental course of language development cross-linguistically and between different dialects of English.

The UK-CDI questionnaire is available as long and short forms and the preliminary norms can be used (Alcock et al., 2017). With this, the new UK-wide CDI has been established and valid UK norms can be used by all interested parties and stakeholders, for example researchers and health services.

So far, no information on predictability of language development is currently available for British English based on the UK-CDI. The current project addressed this gap in the research. The next chapter critically reviews the current literature on the predictive validity of CDIs in languages other than British English or non-normed British CDIs (Oxford CDI, see Hamilton et al., 2000).

3. Literature review of the predictive validity of CDI: Words and Gestures

In order to establish the predictive validity of the UK-CDI for this PhD research the literature of previous research investigating the predictability of CDI: Words and Gestures questionnaires will be reviewed in this chapter. The studies examine the continuity of language on a group level as well as the ability to predict language delay on an individual level. In order to predict later language ability, developmental research investigated the relationship of earlier skills to later acquired language skills (Eriksson, 2001), for example the relation of early vocabulary size to later verbal IQ (Corkum & Dunham, 1996; Pérez-Pereira & Resches, 2011), syntax and morphology (Sachse, Saracino et al., 2007; Sachse, Pecha, & von Suchodoletz, 2007) as well as literacy (Lee, 2011).

Some research suggests good predictive validity from early language skills (e.g., Baumwell & Tamis, 1997; Fenson et al., 1994; Fenson et al., 2007), whilst other research found mixed results (P. Lyytinen et al., 1996) or no relationships between early and later skills (Guiberson, 2008).

These results show that if and how future language outcomes can be predicted from infancy is still an open question with need for agreement. The question of whether language outcomes (e.g. language delay) can be predicted from early stages of communicative development is important as early intervention may yield better long-term outcomes (e.g., Buschmann et al., 2015).

3.1. Background

Predictive value has been established for American English short (Can, Ginsburg-Block, Golinkoff, & Hirsh-Pasek, 2013) and long forms (Feldman et al., 2005) respectively. Furthermore, studies into the predictive validity of CDIs have taken different formats, for example, for non-clinical or clinical groups of participants, for preterm children (Pérez-Pereira, Fernández, Gómez-Taibo, & Resches, 2014; Sansavini et al., 2011), children with low birth weight (Stolt et al., 2009; Stolt et al., 2016), children with cochlear implants (Castellanos, Pisoni, Kronenberger, & Beer, 2016; Nicholas & Geers, 2008), children at risk for Autism (Bopp & Pat, 2011; Miniscalco, Rudling, Råstam, Gillberg, & Åsberg, 2014; Veness, Prior, Eadie, Bavin, & Reilly, 2014), children with developmental disabilities (Yoder et al., 1998) and Down Syndrome (Deckers, Van Zaalen, Mens, Van Balkom, & Verhoeven, 2016) as well as bilingual children (Glennen, 2007; Mancilla-Martinez & Vagh, 2013) and twins (Hayiou-Thomas, Dale, & Plomin, 2012).

As children's language ability grows over time, different domains of language can be tested as they get older. Initially, only communicative gestures, receptive and expressive vocabulary can be recorded, for example, by the CDI: Words and Gestures checklist. Later on, when other skills are acquired in terms of fully-formed utterances on all levels of receptive and expressive language (syntax, pragmatics, semantics, morphology, phonology and phonetics) and literacy, these can be assessed using in part CDI: Words and Sentences for toddlers and CDI-III, or standardised tests.

3.2. Selection criteria

The following review only looked at studies investigating the predictive value of the Infant CDI questionnaires (CDI: Words and Gestures, CDI:WG) to assess the usefulness of early communicative skills as a predictor for future language ability. There were also predictive validity studies using precursor versions (e.g., for highly structered interview checklist see Bates et al., 1988; for ELI see Bornstein & Haynes, 1998; for CDI-WORDS Short Form see Corkum & Dunham, 1996) of the commonly used CDI format (Fenson et al., 1994).

The current project investigated primarily if predictions could be made from the younger infant ages to later language outcomes. Therefore, the literature which used CDI: Words and Sentences for toddlers to investigate the development of typically developing children longitudinally (e.g., Bleses et al., 2016; Dionne et al., 2003; Henrichs et al., 2011; Lee, 2011; Marchman & Bates, 1994; Pan et al., 2004) was not considered here.

Sometimes CDI questionnaires were also used as follow-up for predicting language from early communication skills or other skills associated with language at the initial assessment, for example when children were tested on experimental tasks (e.g., Igualada, Bosch, & Prieto, 2015; Kaduk et al., 2016; Paavola, Kunnari, & Moilanen, 2005; Sundqvist, Nordqvist, Koch, & Heimann, 2016) or after using parent-report questionnaires other than CDIs in regards to different areas, for example language, child temperament, behaviour problems or maternal factors (Longobardi, Rossi-Arnaud, & Spataro, 2011; Prior et al., 2008). This literature was not considered here either as the main focus of interest is on CDIs:WG as a predictor for future language. Details on how the literature search was conducted is given in Appendix 1.

In sum, this review focused on very early communication and the ability to predict future skills from early parent-report. Studies which investigated the predictive value of the CDI:WG, but had a different main aim, for example evaluating the predictive value of a computerized comprehension task in comparison to the CDI:WG (Friend, Schmitt, & Simpson, 2012) were accepted into this review as long as they used the CDI:WG and looked at least at one language measure at follow-up.

3.3. Study characteristics

As the main objective of this project is to assess the predictive validity of the UK-CDI:WG with a typical population, this review only included non-clinical mother and child dyads. However, as a considerable proportion of the typical population has a primary language disorder which is not caused by any other medical condition (around 8% see Norbury et al., 2016), studies investigating language-related deficits were also included, for example families with a history of speech and language impairments (Korpilahti et al., 2016; Spitz, Tallal, Flax, & Benasich, 1997), dyslexia

(Unhjem et al., 2015) or preterm births (Rose, Feldman, & Jankowski, 2009). Studies investigating children with disabilities, children with autism, children with Down Syndrome and bilingual children were excluded (Glennen, 2007). Studies with monolingual children were included with some children being exposed to another language some of the time (Fenson et al., 1994; Fenson et al., 2007; Guiberson, 2008; Thal et al., 1997) or cohorts with a small percentage of bilinguals (7%) in which the number of hours (Rose et al., 2009) were not defined were also included; however, studies with a focus on bilingual children were excluded. These exclusions were followed because the CDI questionnaire is neither tailored to bilingual children with a high exposure to the additional language nor to those with medical problems (see exclusion criteria in the CDI manual, (Fenson et al., 1994; Fenson et al., 2007). Furthermore, Cattani (2014) found that bilingual children had to be exposed to English at least 60% of the time in order to yield results like their monolingual peers on the Oxford CDI (Hamilton et al., 2000). In sum, the aim of this chapter was to review the literature of children with typical monolingual language profiles.

Traditionally, the CDI:WG was used up to 16 months of age; however, since Fenson et al. (2007) updated their norms, they published expanded norms up to 18 months for this instrument. For the purpose of this chapter, studies with the 8- to 18-month age range were taken into account using the CDI:WG. There were also studies investigating the predictability of 18-month-olds' future language using the CDI: Words and Sentences (CDI:WS) (Marschik, Einspieler, Garzarolli, & Prechtl, 2007); however, these studies were not taken into account as we were particularly interested in the CDI:WG and younger age groups. We were only interested in the CDI:WG with its categories (i.e. Phrases Understood, gestures, words understood and said) and not the CDI:WS which investigates more advanced linguistic skills. This was because the current project investigated the predictive validity of the UK-CDI:WG and the literature reviewed here should help to compare results.

Peer-reviewed articles and books were used; however, conference abstracts (e.g., P. Lyytinen & Poikkeus, 1996; Thal, Swaine, Harrison, & Matt, 1998) were not included due to insufficient information given in supplements. There were also articles written in other languages than English or German which could not be included (e.g., Diaz, Fonseca, Bohorquez, Guecha, & Sellabona, 2011; Garcia, Arratibel, Barrena, & Ezeizabarrena, 2008; Serrat et al., 2010; C. Silva et al., 2017).

This comparison of the literature included 26 studies which ranged from the first predictive validity reports in the US norm study (Fenson et al., 1994) until 2016. The languages studied were US American English children (N=11), Finnish (N=5), British English (N=3) and other languages (i.e. Australian English, Bengali, Dutch, Norwegian, Swedish, Mexican Spanish, German). The age for the initial assessment ranged from 9 to 18 months and the age at follow-up ranged from 12 months up to over 8 years of age. Participant numbers also varied widely from 8 (Cochet & Byrne, 2016) to 2156 (Feldman et al., 2000). The samples were diverse with regards socio-economic status (SES). Some studies used heterogeneous SES groups (e.g., Feldman et al., 2000; Friend et al., 2012; Hsu & Iyer, 2016; Laakso et al., 1999), whilst others investigated low (Fish & Pinkerman, 2003; Guiberson, 2008; Hamadani et al., 2010; Kreisman, 2012) or middle to high SES families (Baumwell & Tamis, 1997; Duff, Reen, Plunkett, & Nation, 2015; Fenson et al., 1994). Research design also differed across studies as most studies followed one cohort, but others compared groups over time, for example, children at risk for dyslexia against controls (Laakso et al., 1999; H. Lyytinen et al., 2001; P. Lyytinen et al., 1996; P. Lyytinen, Laasko, Poikkeus, & Rita, 1999; Unhjem et al., 2015) late talkers against controls (Duff et al., 2015), preterm children against controls (Rose et al., 2009), black and white children (Kreisman, 2012) and different reporters (Bornstein et al., 2006). An overview of the studies' characteristics is given in Table 1, for a more comprehensive list including studies' aims see Appendix 2.

Table 1. *Overview of studies's characteristics used in literature review*

Author(s)	Year	Language	Type of participants	Type of CDI (Long/Short Form = SF/LF)	Norms
Baumwell et al.	1997	American English	middle to upper middle class households	CDI:WG (LF) (interview technique used)	no
Bavin et al.	2008	English (Australia)	SES diverse sample (low SES underrepresented)	CDI:WG (LF, parts of Gesture scale), CDI:WS (LF)	no
Bornstein et al.	2006	Dutch in Belgium	first-born, monolinguals, normally developing and term birth children	CDI:WG (LF), CDI:WS (LF)	yes
Cochet & Byrne	2016	British English	no info	Oxford CDI for all age groups (no gesture scale)	no
Duff, Nation, Plunkett & Bishop	2015	British English	31% late talkers at 18 months	Oxford CDI (LF) (no onomatopoeia, no gesture scale)	no
Duff, Reen, Plunkett & Nation	2015	British English	above average SES, 46% had a first degree relative with language and/or reading difficulty	Oxford CDI (no gesture scale)	no
Fenson et al.	1994	American English	sub-dataset from US CDI norm sample	CDI:WG (LF), CDI:WS (LF)	yes
Fenson et al.	2007	American English	sub-dataset from US CDI norm sample	CDI:WG (LF), CDI:WS (LF)	yes
Feldman et al.	2000	American English	sociodemgraphically diverse sample	CDI:WG (LF), CDI:WS (LF)	yes
Fish & Pinkerman	2003	American English	low-SES, rural	CDI:WG (LF)	yes
Friend et al.	2012	American English	Predominantly middle class, demographically diverse sample	CDI:WG (LF), CDI:WS (LF)	yes (but only up to 18 months)
Guiberson	2008	Spanish in US (Mexican immigrants)	low income, low SES families	CDI:WG (LF), CDI:WS (LF)	yes (standardized for Mexico)
Hamadani et al.	2010	Bengali	poor, rural	CDI:WG (SF)	no
Hsu & Iyer	2016	American English	representative sample	CDI:WG (LF)	yes

Korpilahtia et al.	2016	Finnish	population-based sample	CDI:WG (LF), CDI:WS (LF)	yes
Kreisman et al.	2012	American English	White and Black low-income families	CDI:WG (SF with gesture scale), CDI:WS (SF)	yes
Laasko et al.	1999	Finnish	representative parental educational levels, diverse parental reading skills from poor to average	CDI:WG (LF), CDI:WS (LF) and sum score with Bayley expressive score	no
Lyytinen et al.	1996	Finnish	SES representative for the Finnish population, half the parents had problems with reading	CDI:WG (LF), CDI:WS (LF)	no
Lyytinen et al.	1999	Finnish	half of them at risk for dyslexia but no differences in terms of Bayley mental (MDI) scores at 24m	CDI:WG (LF), CDI:WS (LF)	no
Lyytinen et al.	2001	Finnish	half at risk for dyslexia, children from similar educational backgrounds	CDI:WG (LF)	no
Rose et al.	2009	American English, small % Spanish	partially preterm	CDI:WG (SF with gesture scale)	yes
Sachse et al.	2007	German	Addresses were taken from births notices in newspaper (61% at 12m, 95% at 18m)	ELFRA-1, ELFRA-2	yes
Thal et al.	1997	American English	Study 1: US norm sample, Study 2: mostly first-borns, ethnically and racially representative of the US, range of socio-economic backgrounds	CDI:WG (LF), CDI:WS (LF)	yes
Thal et al.	2013	American English	population-based sample	CDI:WG (LF), CDI:WS (LF)	yes
Unhjem et al.	2015	Norwegian	at risk for dyslexia, typical, within typical range for cognitive ability, no neurological disabilities,	CDI:WG (LF), CDI:WS (LF)	yes
Westerlund et al.	2006	Swedish	Swedish primary language, healthy children	CDI:WG (SF)	yes

The correlations and variance contributed to later language abilities were different depending on CDI subscales. See Table 2 for an overview of the content of the original CDI:WG (Fenson et al., 1994) subscales for gestures, phrases (in the following referred to as Phrases Understood) and the vocabulary checklist (total vocabulary checklist score reported separately for production and comprehension).

Table 2. *Overview of the original CDI:WG categories including number of questions and answer types*

Original CDI:WG categories	Number of Questions	Type of Answer
Gestures		
First Communicative Gestures	12	Not yet, sometimes, Often
Games and Routines	6	Yes, No
Actions with Objects	17	Yes, No
Pretending to be a Parent	13	Yes, No
Imitating other Adult Actions	15	Yes, No
Pretend Objects	1	Yes, No
Phrases		
(e.g. Are you hungry)	28	understands
Vocabulary checklist		
Sound Effects and Animal Sounds	12	understands, understands and says
Animals Names	36	understands, understands and says
Vehicles	9	understands, understands and says
Toys	8	understands, understands and says
Food and Drink	30	understands, understands and says
Clothing	19	understands, understands and says
Body Parts	20	understands, understands and says
Furniture and Rooms	24	understands, understands and says
Small Household Items	36	understands, understands and says
Outside Things and Places to go	27	understands, understands and says
People	20	understands, understands and says
Games and Routines	19	understands, understands and says
Action Words	55	understands, understands and says
Words about Time	8	understands, understands and says
Descriptive Words	37	understands, understands and says
Pronouns	11	understands, understands and says
Question Words	6	understands, understands and says
Prepositions and Locations	11	understands, understands and says
Quantifiers	8	understands, understands and says

3.4. CDI: WG – Gestures

Predictive validity of the gesture subscale was examined in several studies as gestures serve as an early means of communication before children use verbal communication (e.g., Acredolo & Goodwyn, 1988; Kuhn et al., 2014). Gesture scores have shown to be strongly related with future language and several studies have shown that gestures predict future language abilities (Iverson & Goldin-Meadow, 2005; Rowe, Özçalişkan, & Goldin-Meadow, 2008; Rowe, Raudenbush, & Goldin-Meadow, 2012).

3.4.1. Are gestures associated with later gestures?

Only three studies tested the predictive validity of early gestures on later gestures (see Table 3), but all showed moderate effect sizes. These held for samples with diverse SES (Fenson et al., 2007) and when the covariates age, gender, birth order and SES was controlled for. Most stability was found within subscales (e.g. when comparing time 1 and time 2 communicative gestures, see Bavin et al., 2008). All three studies examined English speaking children from diverse SES backgrounds (Australian English, Bavin et al., 2008; American English, Fenson et al., 1994; Fenson et al., 2007).

Table 3. *Relationship between gestures and later gestures using the CDI:WG measure*

Authors (date)	N	Predictor measures	Age at predictor	Outcome measures	Age at outcome	Analyses	Covariates	Effect size
Fenson et al. (1994)	62	CDI: WG	Range 8.03 – 10.97 months	CDI: WG	Range 14.93 – 18.63 months	Correlation		r = 0.43 (moderate)
						Regression	Age, Gender, Birth Order, SES	R2 = 0.16 (moderate)
Fenson et al. (2007)	62	CDI: WG	Range 8-10 months	CDI: WG	Range 14 – 17 months	Correlations		r = 0.44 (moderate)
Bavin et al. (2008)	1467	CDI:WG	8 months	CDI: WG	12 months	Correlations		Most predicted associations moderate

3.4.2. Are gestures associated with later comprehension?

Several studies examined the association between infant gesture knowledge and later comprehension, see Table 4. Even though, the types of gestures (Communicative Gestures subscale, other individual CDI gesture subscales (e.g. Actions with Objects), symbolic or representational gestures) were diverse as well as the samples in terms of birth status, ethnic background and family risk status of dyslexia, the results showed a consistent picture despite not controlling for the effect of sample diversity in most studies.

Significant and moderate correlations were found if gestures and comprehension were both tested during infancy of which Actions with Objects (mostly later appearing gestures) contributed the most unique variance to later comprehension (Bavin et al., 2008); however, if all later appearing (symbolic) gestures were correlated with comprehension the associations were weak which may be due to a lower variability in gesture scores during infancy (P. Lyytinen et al., 1996). This was found as part of the Jyväskylä Longitudinal Study of Dyslexia which was conducted in Central Finland and followed-up around 200 children, half of which were at family risk for dyslexia. Furthermore, children with low comprehension scores at 16 months showed the lowest gesture scores several months earlier compared to typically developing children and late producers (Thal, Marchman, & Tomblin, 2013) in a study which followed typically developing American English children's language development from 10 months to 7 years. The aim was to describe language development trajectories and examine factors which might help to detect late-talking at 16 months or earlier. The strength of association decreased with an increasing gap between testing times. Weak correlations were found between infant gestures and comprehension at 3 years in a study investigating Black-White inequality in the US (Kreisman, 2012) and in another study using a partially preterm (31%) and primarily American English-speaking monolingual sample (89.8% monolingual English speakers, 7% bilingual Spanish, 3.2% solely Spanish) (Rose et al., 2009).

Furthermore, no significant relationships were found at 4 and 5 years of age in a study with US American children (Fish & Pinkerman, 2003).

Apart from the time gap between testing times, the instruments employed for assessing comprehension may also play an important role in terms of the strengths of association between gestures and later comprehension. This is because for most languages the CDI vocabulary comprehension subscale was only included in the infant form (up to approximately 18 months), hence other measures of general language comprehension were usually used from that age onwards (see Table 4). It is unclear from the current data if early gestures were more strongly associated with later word comprehension than with general language comprehension and whether different types of gestures influenced word or general language comprehension in different ways and at different time points during development.

Table 4. *Relationship between gestures (using the CDI:WG measure) and later comprehension*

Authors (date)	N	Predictor measures	Age at predictor	Outcome measures	Age at outcome	Analyses	Covariates	Effect size
Bavin et al. (2008)	1467	CDI:WG (individual subscales of CDI gestures)	8 months	CDI:WG	12 months	Correlations		All predicted associations moderate Cohen's d = insufficient info
						Linear regression		R2 = .22 (moderate) Cohen's d = insufficient info
Fish and Pinkerman (2003)	98	CDI:WG (Communicative Gestures)	15 months	Preschool Language Scale (PLS-3)	4 and 5;4 years	Correlations		4 years: r = .09 (none) Cohen's d = .19 (no effect) 5;4 years: r = .19 (weak) (the results were not significant for both correlations) Cohen's d = .39 (small effect)
Kreisman (2012)	1458	CDI:WG (Communicative Gestures) [1]	14 months	Peabody Picture Vocabulary Test (PPVT)	36 months	Correlations		r = .18 (weak) Cohen's d = insufficient info
Laasko (1999)	111	CDI:WG (Gesture Total)	14 months	Reynell Developmental Language Scales (RDLS)	18 months	Hierarchical regression	Predictor entered as first covariate	R2 = .12 (weak to moderate) Cohen's d = insufficient info
Lyytinen et al. (1996)[2]	94	CDI:WG (Symbolic (later appearing) gestures)[3]	14 months	Reynell Developmental Language Scales (RDLS)	18 months	Correlations		r = .20 (weak) Cohen's d = .41 (small effect)

[1] Communicative Gestures form part of the early appearing gestures (see Fenson et al., 1994).
[2] The study by Lyytinen et al. (1996) and Laasko et al. (1999) form part of the Jyväskylä Longitudinal Study of Dyslexia which was conducted in Central Finland and followed-up around 200 children, half of which were at family risk for dyslexia.
[3] Symbolic gestures (i.e. actions with objects, pretending to be a parent, imitating other adult actions, pretending objects) exclude the communicative gestures and games and routines subscale which are among the first gestures learnt.

Rose et al. (2009)	182 (pre-term = 56), full-term = 126)	CDI:WG (Communicative Gestures)	12 months	PPVT	36 months	Correlations		r = .28 (weak to moderate) Cohen's d = insufficient info
						Hierarchical regression	Step 1: birth status (preterm vs full-term birth), Step 2: Infant language at 12 months (comprehension, production, Communicative Gestures)	Infant language as a whole made significant contributions to later language, but Gestures did not add any unique variance Cohen's d = insufficient info
Thal, Marchman and Tomblin (2013)	1107 (863 typically developing children, 154 late producers with typical comprehension, 90 late comprehenders (≤ 10th for comprehension and production)	CDI:WG (early versus later appearing gestures)[4]	10, 13 and 16 months	CDI: WG	16 months	Three-way ANOVA (IV: typically developing, late producers (≤ 10th percentile), late comprehenders (≤ 10th percentile))		Effect sizes not reported At all ages typically developing children and late producers had significantly higher gesture scores (regardless of gesture type) Cohen's d = insufficient info

[4] Early gestures (communicative gestures and games and routines) and later gestures (actions with objects, pretending to be a parent, imitating other adult actions, pretending objects)

3.4.3. Are gestures associated with later production (and grammar)?

Gestures were associated with later language, but the strength of relationship depended on different factors such as age at testing gesture and production or grammar, the type of gestures (early versus later acquired gestures) and the language area tested at follow up (production versus grammar), see Table 5. If gesture knowledge was collected before the first birthday, the relationship with production during toddlerhood was weak (Bavin et al., 2008). However, gestures collected at around the first birthday or during the second year showed mostly moderate relationships with production during toddlerhood in studies looking at German children from primarily middle to upper middle class backgrounds (Sachse, Saracino et al., 2007) and US American children from families with diverse SES backgrounds (Feldman et al., 2000), except for the data from the Finnsih Jyväskylä Longitudinal Study of Dyslexia (Laakso et al., 1999; P. Lyytinen et al., 1996). Lyytinen et al. (1996) used later acquired gestures which showed a weak association with later language. More recent research found that early acquired gestures rather than later acquired gestures contributed the highest variance to later language (Bavin et al., 2008). Laasko et al. (1999) created a composite score of production and grammar at 18 months and found that gestures at 14 months could only weakly predict this composite score. Other research reported here also found that the relationship between gestures and grammar was less strong with relationships varying between weak to moderate in strengths. Furthermore, infant gestures showed no significant relationships with expressive language tests between 3 and 5 years of age. When predicting language status (late talking versus early talking) from infant gestures, studies found that language delay could be predicted up to 28 months albeit only weak effect sizes in studies investigating US American children using population-based samples (Thal et al., 1997; Thal et al., 2013). Early talkers showed better gesture knowledge during infancy compared late or middle range talkers. They also had significantly better gesture skills than those children who were classified as early talkers at 13 months but regressed towards the mean by 20 months. However, even though significant group differences existed, early talking status at 20 months could not be significantly predicted from infant gestures (Thal et al., 1997; Thal et al., 2013).

Table 5. *Relationship between gestures (using the CDI:WG measure) and later production and/or grammar*

Authors (date)	N	Predictor measures	Age at predictor	Outcome measures	Age at outcome	Analyses	Covariates	Effect size
Thal, Marchman and Tomblin (2013)	1107 (863 typically developing children, 154 late producers with typical comprehension, 90 late comprehenders (≤ 10th for comprehension and production)	CDI:WG (early versus later appearing gestures)	10, 13 and 16 months	CDI: WG	16 months	Three-way ANOVA (IV: typically developing, late producers (≤ 10th percentile), late comprehenders (≤ 10th percentile) at 16 months)		Effect sizes not reported At all ages, late producers had significantly better gesture skills than late comprehenders but worse skills than typically developing children regardless of gesture type
		CDI:WG (Gesture Total)	16 months	CDI:WS	28 months	Correlations		Production: r = .40 (moderate); Sentence Complexity: r =.35, (moderate)
		CDI:WG (Gesture Total)	16 months	CDI:WS (Production and Sentence Complexity)	28 months	Logistic regression (outcome variable: language delay ≤ 10th percentile on CDI:WS vocabulary production AND/OR sentence complexity score at 28 months)		Somer's D = .12 (weak)
Feldman et al. (2000)	2156	CDI:WG (Gesture Total)	12 months	CDI:WS	24 months	Correlations		Production: r = .34 (moderate), Grammar: weak to moderate
Sachse et al. (2007)	149	ELFRA-1[5] (Gesture Total)	12 months	ELFRA-2[6]	24 months	Correlations		Production: rs = .42 (moderate)

[5] ELFRA questionnaires (Elternfragebogen für die Früherkennung von Risikokindern) are CDI adaptions in a broad sense as they do not describe the typical course of language development; however, instead aim to identify children at risk for language problems at 1;0 and 2;0 years (Grimm & Doil, 2006).
ELFRA-1 consists of a vocabulary checklist (comprehension, production), gestures and motor behaviour
[6] ELFRA-2 contains a relatively short vocabulary checklist (260 words for production only) and questions about grammar (morphology and syntax)

Lyytinen et al. (1996)	94	CDI:WG (Symbolic gestures)	14 months	CDI:WS	18 months	Correlations		Production: r = .28 (weak) Grammar (use of suffixes): weak and not significant
		CDI:WG (Symbolic gestures)	14 months	Reynell Developmental Language Scales (RDLS)	18 months	Correlations		r =.19 (weak)
Laasko et al. (1999)	111	CDI:WG (Gesture Total)	14 months	Composite score: CDI:WS (production & maximum sentence length (MSL)) and Bayley expressive score (Bayley, 1993)	18 months	Hierarchical regression	Predictor entered as first covariate	R2 = .10 (weak)
Bavin et al. (2008)	1467	CDI:WG (individual subscales of CDI gestures)	8 and 12 months	CDI:WG (12 months) and CDI:WS (24 months) (Production)	12 and 24 months	Correlations		Predicted associations between 8 and 12 months as well as between 12 and 24 months were mostly moderate, all correlations between 8 and 24 months were weak
		CDI:WG	8 months	CDI:WG (Production)	12 months	Linear regression	predictor variables: Communicative Gestures, Games and Routines, Actions with Objects	R2 = .14 (moderate) All predictors were significant and Communicative Gestures contributed the highest unique variance of 2.16%

		CDI:WG	8 months	CDI:WS (Production)	24 months	Linear regression	Same as above	R2 = .05 (weak) Only Communicative Gestures (1.15%) and Games and Routines (.55%) contributed unique variance *(these are early acquired gestures)*
		CDI:WG	12 months	CDI:WS (Production)	24 months	Linear regression	predictor variables: Communicative Gestures, Games and Routines, Actions with Objects, Pretending to be a Parent, Imitating other Adult Actions	R2 = .15 (moderate) All predictors were significant except Imitating other Adult Actions. Communicative Gestures contributed the highest variance of 1.5%
Kreisman (2012)	1458	CDI:WG (Communicative Gestures)	14 months	CDI:WS	24 months	Correlations		Production: r = .37 (moderate) Sentence Complexity: r = .25, (weak)
Rose et al. (2009)	182 (pre-term = 56), full-term = 126)	CDI:WG (Communicative Gestures)	12 months	verbal fluency (age-appropriate modification of the Educational Testing Service (ETS))	36 months	Correlations		r = .15 (weak and not significant)
						Hierarchical regression	Step 1: birth status (preterm vs full-term birth), Step 2: Infant language at 12 months (comprehension, production,	Infant language as a whole made significant contributions to later language, but Gestures did not add any unique variance

							Communicative Gestures)	
Fish and Pinkerman (2003)	98	CDI:WG (Communicative Gestures)	15 months	PLS-3	4 and 5;4 years	Correlations		4 years: r = .17 (weak) 5;4 years: r = .21 (weak) Neither correlation was statistically significant
Thal et al. (1997)	217	CDI:WG (Gesture Total)	Range between 10 and 16 months (Mean = 13.45)	CDI:WS (Production)	Range between 16 and 25 months (Mean = 20.15)	ANOVA (IV: early talker (≥ 90th percentile), middle range talker, late talker (≤ 10th percentile) at 20 months[7])		No effect sizes reported, ANOVA was significant and post hoc Tukey tests revealed that early talkers (mean = 43.74) used significantly more gestures at time 1 than the other two groups, middle range talkers (mean = 31.25) used significantly more gestures than late talkers (mean = 23.33)
						T-tests with Bonferroni adjustment (IV: early talkers who remained early talkers, early talkers who regressed towards mean at 20 months)		Eta squared = .19 (large) Children who remained early talkers used significantly more gestures at 13 months than those early talkers who regressed towards the mean at follow-up (p < .004; early at both times mean = 33.75, typical at time 2 mean = 42.69)
						Regression (Outcome variable: late-talker status at 20 months)	time 1 age, time 2 age, SES, word comprehension, production,	R2 = .06 (weak)

[7] Gender-specific norms were used to establish language status

			percentage of all words understood that were also produced	
		Regression (Outcome variable: early-talker status at 20 months)	time 1 age, time 2 age, SES, word comprehension, production, percentage of all words understood that were also produced	Gesture accounted for no unique variance

3.4.4. Are gestures associated with later total language scores or risk for language impairment?

To add to the previous section, I now describe the relationships between gestures and later total language skills or risk for language impairment of which both outcome measures were made of production as well as comprehension abilities. This is important as we found that the relationships between gestures and production were generally of moderate strengths, whilst the associations between gestures and comprehension were somewhat less strong. As expected from the previous studies, the current data showed weak relationships between gestures and total language during toddlerhood and no significant relationships if tested at 4 or 5 years, see Table 6. Furthermore, similarly to predicting late-talking at 28 months (Thal et al., 2013), children at risk for language impairment at 3 years showed lower gesture skills at 15 months but this relationship was also weak in a study using a large US American sample (N = 1064) (Hsu & Iyer, 2016). However, children at risk for language impairment at 4;6 years did not differ in terms of gesture scores at 15 months. Whilst gestures at 15 months could not directly predict risk for language impairment at 3 or 4;6 years of age, there was an indirect effect for gestures as language impairment was mediated by production skills at 15 months (Hsu & Iyer, 2016).

Table 6. *Relationship between gestures (using the CDI:WG measure) and later total language or risk for language impairment*

Authors (date)	N	Predictor measures	Age at predictor	Outcome measures	Age at outcome	Analyses	Covariates	Effect size
Kreisman (2012)	1458	CDI:WG (Communicative Gestures)	14 months	Bayley language factor (12 selected items from the receptive and expressive subscales of the Bayley Scales of Infant Development, 2nd edition)	24 months	Correlations		r = .20 (weak)
Fish and Pinkerman (2003)	98	CDI:WG (Communicative Gestures)	15 months	PLS-3 (total language score)	4 and 5;4 years	Correlations		4 years: r = .17 (weak) 5;4 years: r = .22 (weak) Neither correlation was statistically significant
Hsu and Iyer (2016)	1064	CDI:WG (Communicative Gestures)	15 months	At-risk status: RDLS (total score)	3 years	T-test (children at risk for language impairment at 3 years versus children not at risk)		Children who were at risk at 3 years had produced significantly fewer communicative gestures, the effect was small (eta squared = .008)
				At-risk status: PLS-3 (total language score)	4;6 years	T-test (children at risk for language impairment at 4;6 years versus children not at risk)		Effect size not reported, t-tests was not significant
		CDI:WG (Communicative Gestures)	15 months	3 years: RDLS (total score), 4;6 years: PLS-3 (total language score)	composite of at-risk status at 3 and 4;6 years	Structural Equation Model (SEM)	Mediator: CDI:WG production	Gestures at 15 months did not directly predict later LI risk but the influence of early gestures on later risk for language impairment was mediated by production skills at 15 months

3.4.5. Summary

Overall, gestures were more strongly associated with later production than grammar, comprehension or a composite of production and comprehension. The most robust associations were generally found between infant gestures and language scores up to around 2 years of age (moderate strengths) but relationships ceased to exist from around 3 years onwards. Gestures could predict language delay at 28 months; however, the effect was weak. Later language impairment could not be predicted from infant gesture scores alone but in conjunction with early production skills could help to predict at risk status for language impairments up to three years later (at 3 and 4;6 years of age).

3.5. CDI:WG – Phrases Understood

3.5.1. Are Phrases Understood scores associated with later language?

Out of the 26 studies reviewed here, only two US American studies reported longitudinal predictions between the subscale Phrases Understood and later language, see Table 7. As reported for gestures, the relationship was stronger between Phrases Understood and word production than with grammar at 24 months. In addition, the strength of relationship decreased with time from moderate at 24 months for word production to weak when tested at 4 and 5;4 years using the PLS-3. At 4 years, the effect of Phrases Understood remained significant even after controlling for gender, articulation and behaviour problems at 4 years and when using diverse samples. Phrases Understood has more in common with word comprehension rather than word production, thus it was somewhat surprising that Phrases Understood correlated better with later measures of expressive rather than receptive language particularly because other categories showed stronger long-term correlations within the same domain.

Table 7. *Relationship between Phrases Understood (using the CDI:WG measure) and later language*

Authors (date)	N	Predictor measures	Age at predictor	Outcome measures	Age at outcome	Analyses	Covariates	Effect size
Feldman et al. (2000)	2156	CDI:WG	12 months	CDI:WS	24 months	Correlations		Production: r = .31 (moderate) Grammar: weak
Fish and Pinkerman (2003)	98	CDI:WG (Communicative Gestures)	15 months	PLS-3 (expressive communication and total language score)	4 years	Correlations		All weak associations
				PLS-3 (expressive communication and total language score)	5;4 years	Correlations		expressive communication: r = .03 (no relationship) total language: r = .05 (no relationship)
				PLS-3 (auditory comprehension)	4 and 5;4 years	Correlations		4 years: r = .17 (weak and not significant), 5;4 years: r = .05 (no relationship)
				PLS-3 (expressive language)	4 years	hierarchical regression	step 1 articulation problems at 4 years, uncooperative behaviour at 4 years and child's sex, step 2 CDI Phrases Understood at 15 months, child's lack of initiative at 4 years and child's behavioural problems at 4 years, step 3 maternal facilitation at 9 months and maternal over-control at 4 years	Step 2: R^2 = .19 (moderate) (of which Phrases Understood and child's behavioural problems were significant after all predictors had been added)
				PLS-3 (auditory comprehension)	4 years	hierarchical regression	step 1 articulation problems at 4 years, uncooperative behaviour at 4 years and child's sex, step 2 number of books at 4 years, mother smokes, step 3 CDI Phrases Understood at 15 months, child's lack of initiative at 4 years and child's behavioural problems at 4 years, step 4 maternal facilitation at 9 months and maternal over-control at 4 years	Step 2: R^2 = .19 (moderate) (of which all three predictors were significant after all predictors had been added)

3.6. CDI:WG - Comprehension

3.6.1. Is early vocabulary comprehension associated with later comprehension?

Parent report of early word comprehension may yield less accurate results than those of early word production (as suggested above, see *Critical assessment of CDIs*). An important factor for accurate reports was parental verbal sensitivity which has also been suggested to advance children's cognitive and linguistic abilities and was particularly important for those children with low initial comprehension scores in a study of monolingual American English from middle to upper middle class backgrounds (Baumwell & Tamis, 1997). Hence, when controlling for maternal sensitivity it was not surprising that associations between comprehension scores at 9 and 13 months were weak (Baumwell & Tamis, 1997). However, without using this control, the associations between infant word comprehension and later comprehension (up to around 20 months of age) were moderate to large in strength using the CDI:WG at both time points when looking at the original US American MB-CDI data (Fenson et al., 1994; Fenson et al., 2007), data from different respondents (mothers, fathers and other person close to the child) completing the CDI for Dutch children (Bornstein et al., 2006), CDIs of Norwegian monolingual control children in a study investigating the continuity of langue development during the second year in children at risk of dyslexia and control children (Unhjem et al., 2015) and data from a large sample (N= 1429) of children from a poor rural district in Bangladesh (Hamadani et al., 2010), see Table 8. An exception was the study by Lyytinen et al. (1996) who found weak correlations when using the CDI:WG at 14 months and the RDLS at 18 months as part of the Finnish Jyväskylä Longitudinal Study of Dyslexia; however, after including more children in the study they also reported moderate associations (P. Lyytinen et al., 1999).

Further support of the continuity of comprehension scores was reported by Bornstein et al. (2006) who found no significant differences in the correlations between different respondent types (e.g. mother, father). In addition, even after controlling for age at time 1, gender, birth order and SES, Fenson et al. (1994) showed that comprehension at 8 months explained a moderate amount in comprehension scores at 16 months and after controlling for prematurity, Rose et al. (2009) reported moderate correlations between 12 and 36 months. Again, the strength of correlations decreased over time and was weak when correlating infant comprehension scores with comprehension assessed at 4 or 5 years of age (Fish & Pinkerman, 2003).

For some studies, the correlations were already weak between infant comprehension and comprehension at 3 years when looking at US American English children from Caucasian and African American ethnic groups (Kreisman, 2012) and at a population-based sample of monolingual Finnish children (Korpilahti et al., 2016). This was probably because former low comprehenders had moved up into the typical range on standardised comprehension tests at kindergarten age when on some tests, they did not significantly differ from typical children anymore (Thal et al., 2013). However, on some tests low comprehenders still lacked significantly behind typical children at 4 and 5 years (Thal et al., 2013).

Whilst low ability children caught up over time, high ability children remained more stable (Thal et al., 2013). We would therefore expect stronger correlations for high ability children compared to low ability children over time. This may explain the strong associations between comprehension scores in the study by Cochet and Byrne (2016) who used a small British English sample (N= 8) between 11 and 41 months and investigated the developmental continuities between early social and communicative (including gesture, language) abilities and language development. Possibly due to the high variation in age at both time points, no raw scores were reported and no information was given about the children's background, thus whilst we cannot be certain, it may be that these children had high language scores from the outset of language development which could explain the strong correlation between time 1 and time 2. It was mentioned that one of the children was excluded due to reaching ceiling on the Oxford CDI. Due to the small sample size used it is unclear if a replication of the study with a more representative sample would yield the same strengths of long-term associations between comprehension scores. This was a laboratory study which often include Caucasian children from middle class backgrounds, hence this heightens the possibility that these children had average or above average language skills overall. Furthermore, the mean age of the children at time 1 (21;8 months) was a lot older compared to the other studies presented here; as language has shown to be more stable during toddlerhood (e.g., Feldman et al., 2005; Pérez-Pereira & Resches, 2011; Ullrich & von Suchodoletz, 2011) stronger relationships were expected at this age. This study also used the Oxford CDI for children up to 41 months which may not be appropriate. Even though, the authors reported that the effect for age on the CDI was not significant and therefore included all children, the questionnaire was initially made for children up to 25 months and thus did not aim to distinguish between ability levels at 41 months of age.

Overall, associations between comprehension scores were generally moderate in strength if assessed during infancy at both time points. The relationship became weak when analysing the association between infant comprehension and comprehension scores at kindergarten age but remained significant even after controlling for many different factors. This decrease in the strength of the relationship was probably because most of the children with low comprehension at the outset of language development managed to catch up with their peers at 4 to 5 years of age (Thal et al., 2013).

Table 8. *Relationship between comprehension (using the CDI:WG measure) and later comprehension*

Authors (date)	N	Predictor measures	Age at predictor	Outcome measures	Age at outcome	Analyses	Covariates	Effect size
Baumwell et al. (1997)	40	CDI:WG	9 months	CDI:WG	13 months	Regression	mothers' sensitivity at 9 and 13 months	R2 = .10 (small)
						Regression	Moderator: mothers' sensitivity at 9 months	R2 = .08 (small) the effect of maternal sensitivity on later comprehension was stronger for children with lower comprehension scores at 9 months
Thal et al. (2013)	1107 (863 typically developing children, 154 late producers with typical comprehension, 90 late comprehenders (≤ 10th for comprehension and production)	CDI:WG	10 and 13, 16 months	CDI:WG	16 months	Three-way ANOVA (IV: typically developing, late producers (≤ 10th percentile), late comprehenders (≤ 10th percentile) at 16 months)		Effect size not reported late comprehenders at 16 months understood significantly fewer words than their typically developing peers and late producers at all ages
	90 (typical, late producer and late comprehender group sizes not specified)			CELF-P Receptive, PPVT-R	4 years	Three-way ANOVA (see above)		Effect size not reported All groups were within normal range but typical children had significantly higher scores than the delayed group on the CELF-P, no differences between groups on the PPVT-R
	90 (typical, late producer and late comprehender group sizes not specified)			CELF-P Receptive, PPVT-R	5 years	Three-way ANOVA (separately for both instruments)		Effect size not reported All groups had scores within the typical range; typically developing children had significantly higher scores on the PPVT-R than the delayed groups, no differences between groups on the CELF-P
Fenson et al. (1994)	62	CDI:WG	8 months	CDI:WG	16 months	Correlation		r = .48 (moderate)

						Multiple Regression	age, gender, birth order, SES	R2 = .22 (moderate)
Fenson et al. (2007)	62	CDI:WG	8-10 months	CDI: WG	14 – 17 months	Correlation		r = .44 (moderate)
Unhjem et al. (2015)	53 (children with familial risk for dyslexia (N= 32) and control children (N= 21)	CDI:WG	12 months	CDI: WG	15 months	Correlations		at risk: .69 (large) control: r = .61 (large)
		CDI:WG	12 months	CDI: WG	18 months	Correlations		at risk: r = .82 (large), control: r = .49 (moderate)
		CDI:WG	15 months	CDI: WG	18 months	Correlations		at risk: r = .79 (large), control: r = .68 (large)
Lyytinen et al. (1999)	94 (half of the parents had problems in reading)	CDI:WG	14 months	RDLS comp	18 months	Correlation		r = .23 (small)
Lyytinen et al. (1999)	171 (half of them at familial risk for dyslexia)	CDI:WG	14 months	RDLS comp	18 months	Correlation		r = .40 (moderate)
Laasko et al. (1999)	111 (parents had average to poor reading skills)	CDI:WG	14 months	RDLS comp	18 months	Correlation		r = .38 (moderate)
Bornstein et al. (2006)	29	CDI:WG	13 months	CDI:WS	20 months	Correlations (with different respondents)		All correlations were large (and were not significantly different from each other)
Rose et al. (2009)	182 (pre-term = 56), full-term = 126)	CDI:WG	12 months	Peabody Picture Vocabulary Test, PPVT	36 months	Correlation	birth status (preterm vs full-term birth)	r = .49 (moderate)
						Hierarchical regression	Step 1: birth status (preterm vs full-term birth), Step 2: Infant language at 12 months (comprehension,	Infant language: R2 = .29 (moderate) of which comprehension added the strongest contributions (after

							production, Communicative Gestures)	adding infant information processing as a third step, only comprehension remained significant for infant language)
Korpilahti et al. (2016)	226	CDI:WG	13 months	RDLS-III comp	36 months	Correlation		r = .18 (weak)
				RDLS-III comp	36 months	Multiple linear regression	CDI:WG production, comprehension at 13 months CDI:WS production at 24 months	Comprehension scores at 13 months accounted for no significant variance; only production at 24 months was significant (R2 = .19)
Kreisman (2012)	1458	CDI:WG	14 months	PPVT	36 months	Correlation		r = .17 (small)
Hamadani et al. (2010)	1429	CDI:WG	12 months	CDI:WG	18 months	Correlation		r = .51 (large)
Cochet and Byrne (2016)	8	Oxford CDI[8]	11 – 33 months	Oxford CDI	21 – 41 months	Correlation		rs = .72 (large)
Fish and Pinkerman (2003)	98	CDI:WG	15 months	PLS-3	4 and 5;4 years	Correlations		4 years: r = .19 (weak) 5;4 years: r = .11 (weak) Neither relationship was statistically significant
					4 years	Hierarchical regression	step 1 articulation problems and uncooperative behaviour, step 2 contextual variables, step 3 words understood at 15 months, child's behaviour problems and lack of initiative at 4 years, step 4 mother-child interaction	Step 3: R2 = .19 (moderate) in the final model all three variables of step 3 were significant but comprehension added the smallest contributions

[8] Oxford CDI is a UK adaption of the CDI:WG which is not standardized for the UK. It does not include a gesture subscale.

3.6.2. Is early vocabulary comprehension associated with later production (and grammar)?

Many studies examined the relationship between infant comprehension and later production (and grammar), see Table 9. The reason for this may be that during early development there was more variability in word comprehension than production as children understood many words before they started to produce them (Fenson et al., 1994). Furthermore, some research suggested that word comprehension was somewhat better than production at predicting later production scores (e.g., Thal et al., 2013).

As part of the *Critical assessment of CDIs* chapter it was indicated that caution should be taken when using early comprehension scores of parents from low SES backgrounds as it has been repeatedly reported that early comprehension scores were higher in groups with low SES compared to high SES status but during toddlerhood language scores were lower in children from low SES backgrounds (Arriaga et al., 1998). Therefore, it could be that the strengths of association between infant word comprehension and later language was weaker for groups with low SES compared to high SES status. However, from the current data SES status showed no differences in the strengths of relationships for groups with low SES status (Fish & Pinkerman, 2003; Kreisman, 2012) compared to the SES diverse samples used in the other studies.

Most studies investigated the relationship between word comprehension between 12-14 months and production (and grammar) at 24 months. Whilst SES status did not yield different results, the sample size seemed to influence the strengths of relationships. All studies with more than 1000 participants showed moderate relationships (Bavin et al., 2008; Feldman et al., 2000; Kreisman, 2012) and all studies with smaller samples (ranging from 21 to 226 participants) showed no significant or weak relationships between comprehension and production (Korpilahti et al., 2016; P. Lyytinen et al., 1996; Sachse, Saracino et al., 2007; Unhjem et al., 2015). A similar picture arose for the association between comprehension and grammar but in all situations the relationship was somewhat weaker than between comprehension and production in the same study (Feldman et al., 2000; Kreisman, 2012; P. Lyytinen et al., 1996). When studying different ability groups over time the language of the lowest ability children was least stable (e.g., Thal et al., 2013), thus in small samples a diversity of abilities lead to weaker correlations compared to bigger samples in which the percentage of typical children (referring to the normal distribution in statistical terms) with more stable language was higher (Thal et al., 1997).

Furthermore, the strengths of association may also be affected by the population used (i.e. children with or without family risk of dyslexia). For example, the study by Unhjem et al. (2015) used typically developing control children and found large correlations between 15 and 18 months between comprehension and production, in contrast Lyytinen et al. (1996) used a sample in which half of the parents had reading difficulties and found small correlations between 14 and 18 months. Whilst it should be considered that the study by Unhjem et al. (2015) used a shorter age gap (1 month) and a composite production score created from the CDI:WG and CDI:WS which may have also influenced the results, the difference in strengths of association was considerable. This notion that different populations differed in the continuity of language is further supported by the findings from Unhjem et al. (2015) who reported that the contributions of infant comprehension to expressive Bayley scores at 24 months was significant and large for children at-risk for dyslexia but not significant for typically developing control children. This showed that there was a stronger continuity in the development of language in the at-risk children and that comprehension predicted later language (as suggested above, see Thal et al., 2013) but only for the at-risk children and not the typical children.

Furthermore, the association between infant comprehension and later production (at 28 months) was moderate in strength if the comprehension scores were collected during the second year (at 16 months) demonstrating a stronger relationship than if growth scores between 10 and 16 months were analysed. This shows that CDI comprehension scores were more stable if tested later during development rather than during the early stages of language development (Thal et al., 2013). In addition, the strengths of relationship between early comprehension and later production decreased when production was tested later than at 24 months of age showing mostly weak to moderate associations when tested at 28 months, 36 months, 4 years and 5 years, except for Cochet and Byrne (2016) who reported large associations possibly due to sampling as discussed above. In addition, the data indicated that stronger correlations were found between CDI instruments compared to different measures used at follow-up in a study of American English children from predominantly middle class families (e.g. spontaneous speech, see Friend et al., 2012).

In addition, the relationship remained stable even after controlling for the child's history of ear infections, family history of learning disabilities, family history of speech and language disorders, parental education and income (Thal et al., 2013). Interestingly, when predicting verbal fluency at 36 months from CDI scores at 12 months, Rose et al. (2009) found that comprehension and production contributed significant variance after controlling for prematurity. However, after adding information processing (of which only tactual-visual cross-modal competence was significant) as a last step in the model, production but not comprehension was still significant. Language comprehension was concurrently correlated with cross-modal competence and when both were included in the model, cross-modal competence was a better predictor for future language than comprehension whilst production remained a significant predictor unaffected by infant information processing skills. These results supported a domain-general view of language showing that language did not only rely on language specific processes but also on other cognitive mechanisms (Rose et al., 2009).

When studying the trajectories of children with different abilities, the language of late, middle and early talkers remained stable between 13 and 20 months, particularly if the *percent of words understood and also produced* was examined at 13 months rather than comprehension. Whilst infant comprehension (1.8%) and production (2.1%) explained some small variance in late-talking status at 20 months, the *percent of words understood and also produced* (3.4%) together with gestures (6%) contributed significant and substantial variance to early talking status in later development (Thal et al., 1997). This showed that the continuity of language was stronger for early talkers than for late talkers. In addition, late comprehenders remained significantly less skilled than typically developing children and late producers in terms of vocabulary at 28 months, grammar at 28 and 36 months as well as on standardised language tests at 4 years but no differences were found at 5 years on standardised tests between groups. The scores of late comprehenders moved up to the typical range on language tests from 4 years. At 5 years, no differences could be detected for late comprehenders and late producers compared to typically developing children on expressive language tests (Thal et al., 2013).

In sum, the relationship between comprehension and later production (and grammar) during infancy and toddlerhood was significant but varied in strength depending on sample size and population used. The relationship was also less strong than for associations within the same category (e.g. early comprehension and later comprehension). Whilst, the association remained stable after controlling for several factors (for example SES and medical history), it became clear that other cognitive skills during infancy were even better at predicting later production than early comprehension (Rose et al., 2009). As reported before, scores during infancy correlated only weakly with scores at kindergarten age and were not significant anymore at that age (Fish & Pinkerman, 2003). This was probably due to the fact that whilst early talkers remained stable over time in language ability, late talkers caught up with their peers at around 4 to 5 years (Thal et al., 2013).

Table 9. *Relationship between comprehension (using the CDI:WG measure) and later production and /or grammar*

Authors (date)	N	Predictor measures	Age at predictor	Outcome measures	Age at outcome	Analyses	Covariates	Effect size
Thal et al. (2013)[9]	1107 (863 typically developing children, 154 late producers with typical comprehension, 90 late comprehenders (≤ 10th percentile for comprehension and production)	CDI:WG	10 and 13, 16 months	CDI:WG	16 months	Three-way ANOVA (IV: typically developing, late producers (≤ 10th percentile), late comprehenders (≤ 10th percentile) at 16 months)		Effect size not reported Late producers had significantly lower scores than typical children but both groups (typical and late producers) scored significantly higher than late comprehenders
		CDI:WG	16 months	CDI:WS production	28 months	Three-way ANOVA		For all three ANOVAS, late producers had significantly lower scores than typical children but both groups scored significantly higher than late comprehenders: amount of variance accounted for was 23% for vocabulary size
		CDI:WG	16 months	CDI:WS	28 months	Correlations		production: r = .49 (moderate) sentence complexity: r = .41 (moderate)
	391 (262 typically developing children, 86 late producers, 43 late comprehenders)	CDI:WG	16 months	CDI:WS (word combinations)	20, 24, 28 and 36 months	Three-way ANOVA		No effect sizes reported late comprehenders scored significantly lower than typical children but did not differ from late producers at 20 and 24 months, the three groups

[9] Children from all groups scored within the normal range at 28 months and at 4 and 5 years

						did not differ from each other at 28 and 36 months (all groups reached ceiling)
CDI:WG	16 months	CDI:WS (complex sentences)	20, 24, 28 and 36 months	Three-way ANOVA		At 20 months: late comprehenders did not significantly differ from the other groups (floor effects) At 24, 28 and 36 months: late comprehenders scored significantly lower than both typical children and late producers
CDI:WG	16 months	CDI:WS (M3L)	20, 24, 28 and 36 months	Three-way ANOVA		At 20, 24 months: late comprehenders scored significantly lower than typical children but did not differ from late producers At 28 and 36 months: late comprehenders scored significantly lower than both typical children and late producers
CDI:WG	16 months	CDI:WS (language delay ≤ 10^{th} percentile for vocabulary production AND/OR sentence complexity score)	28 months	regression	Predictor variables: child's history of ear infections by 16 months, family history of speech and language disorders, family history of learning disabilities, maternal and paternal education, family income, CDI:WG at 16 months comprehension, production, gesture	Unique contribution of comprehension Somer's D = .16 (weak)

		CDI:WG	10 and 16 months	CDI:WS (production)	28 months	Stepwise multiple regression	Predictor variables: growth in scores between 10 and 16 months for comprehension, production and gesture, CDI:WG scores at 16 months for comprehension, production and gesture	Growth in Comprehension scores R2= .01 (small) Comprehension at 16 months R2=.24 (moderate) (Comprehension at 16 months contributed the highest unique variance)
	90 (group sizes not specified)	CDI:WG	16 months	Expressive Vocabulary Test (EVT)[10]	5 years			Late comprehenders did not differ from late producers and typical children
	90 (group sizes not specified)	CDI:WG	16 months	Expressive subscale of the Clinical Evaluationof Language Fundamenta ls Preschool (CELF-P)[11]	4 and 5 years			At 4 years: Late comprehenders scored significantly lower compared to typical children but did not differ from late producers At 5 years: Late comprehenders did not differ from late producers and typical children
Feldman et al. (2000)	2156	CDI:WG	12 months	CDI:WS	24 months	Correlations		production: r = .34 (moderate) grammar: weak to moderate
Bavin et al. (2008)	1467	CDI:WG	12 months	CDI:WS	24 months	Correlation		production: r = .38 (moderate)

[10] Williams (1997)
[11] Wiig, Secord & Semel (1992)

Kreisman (2012)	1458	CDI:WG	14 months	CDI:WS	24 months	Correlations		production r = .42 (moderate) sentence complexity: r = .29 (small)
Sachse et al. (2007)	149	ELFRA-1	12 months	ELFRA-2	24 months	Correlation		production: rs = .35 (weak)
Lyytinen et al. (1996)	94 (half of the parents had problems in reading)	CDI:WG	14 months	CDI:WS	18 months	Correlations		production: r = .25 (small) use of suffixes: r = .20 (small)
				RDLS (expressive language)	18 months	Correlations		r =.10 (no relationship)
Lyytinen et al. (1999)	171	CDI:WG	14 months	CDI:WS (production)	24 months	Hierarchical multiple regression	Predictor variables: maternal education, gender, CDI comprehension at 14 months, CDI production at 14 months, nonsymbolic play, symbolic play	Comprehension R2 = .01 (very weak and not significant) Entire Model R2 = .27 (large)
				CDI:WS (Maximum sentence length)	24 months	Hierarchical multiple regression	Predictor variables: maternal education, gender, CDI comprehension at 14 months, CDI production at 14 months, nonsymbolic play, symbolic play	Comprehension R2 = .01 (very weak and not significant) Entire Model R2 = .27 (large)
				Bayley expressive score	24 months	Hierarchical multiple regression	Predictor variables: maternal education, gender, CDI comprehension at 14 months, CDI production	Comprehension R2 = .00 (very weak and not significant) Entire Model R2 = .27 (moderate)

							at 14 months, nonsymbolic play, symbolic play	
Unhjem et al. (2015)	32 (at familial risk for dyslexia)	CDI:WG	12,15 and 18 months	CDI:WG (15 months) Composite production of CDI:WG & CDI:WS (at 18 months), production composite of CDI:WS & Bayley (24 months)	15, 18 and 24 months	Correlations		All associations moderate to large
	21 (control children)	CDI:WG	12,15 and 18 months	see above	15, 18 and 24 months	Correlations		12 to 24 months: r = .34 (moderate and not significant) 15 to 18 months: r = .68 (large) 15 to 24 months: r = .16 (small) 18 to 24 months: r = .53 (large)[12]
	32 (at familial risk for dyslexia)			Bayley production	24 months	Hierarchical regression	step 1 CDI comprehension at 12 months, CDI production at 12 months; step 2 CDI comprehension 15 months, CDI production at 15 months; step 3 CDI comprehension at 18	Comprehension at 12 months: R2 = .28 (large) Comprehension at 18 months: R2 = .08 (weak) Entire model R2 = .65 (large) most variance was contributed by composite productive vocabulary at 18 months (R2 = .30)

[12] Fisher's z-transformed correlations coefficients showed that the correlations with 24 months production was not significantly different between the at-risk and control groups, all other correlations were significantly different from each other

							months, CDI composite production at 18 months	
	21 (control children)			Bayley production	24 months	Hierarchical regression	step 1 CDI comprehension at 12 months, CDI production at 12 months; step 2 CDI comprehension 15 months, CDI production at 15 months; step 3 CDI comprehension at 18 months, CDI composite production at 18 months	Comprehension at 12 and 18 months explained no significant variance in Bayley production scores at 24 months Entire model R2 = .34 (large) the only significant predictor was CDI production at 12 months (R2 = .34)
Korpilahti et al. (2016)	226	CDI:WG	13 months	CDI:WS	24 months	Correlation		production: r = .27 (weak)
				Renfrew lexical skills	36 months	Correlation		r = .15 (weak)
				Renfrew lexical skills	36 months	Multiple linear regression	CDI:WG production, comprehension at 13 months CDI:WS production at 24 months	Comprehension scores at 13 months accounted for no significant variance; only production at 24 months was significant (R2 = .15)
Rose et al. (2009)	182 (pre-term = 56), full-term = 126)	CDI:WG	12 months	Educational Testing Service test for verbal fluency[13]	36 months	Correlation	birth status (preterm vs full-term birth)	r = .29 (weak)
						Hierarchical regression	Step 1: birth status (preterm vs full-term), Step 2: infant language at 12 months (comprehension,	Infant language: R2 = .12 (small to moderate) and comprehension and production were significant predictors

[13] Singer, Corley, Guiffrida, & Plomin (1984)

							production, Communicative Gestures))	(after adding infant information processing as step 3 only production was significant)
Cochet and Byrne (2016)	8	Oxford CDI[14]	Age range between 11 – 33 months	Oxford CDI	Age range between 21 – 41 months	Correlation		rs = .75 (large)
Friend et al. (2012)[15]	24 (partially preterm and mostly monolingual)	CDI:WG	Age range between 16;2 - 21;4	CDI:WS	Age range between 24 - 41	Correlations		Production: r = .50 (moderate to large) MLU[16]: r = .61 (large)
			Age range between 16;2 - 21;4	Spontaneous vocabulary (different words or unique word roots)	Age range between 24 - 41	Correlation		r = .40 (moderate but relationship was not statistically significant)
Fish and Pinkerman (2003)	98	CDI:WG	15 months	Expressive PLS-3	4 and 5;4 years	Correlations		4 years: r = .15 (weak) 5;4 years: r = .14 (weak) Neither correlation was statistically significant
Thal et al. (1997)	217	CDI:WG (Comprehension)	Range between 10 and 16 months (Mean = 13.45)	CDI:WS (Production)	Range between 16 and 25 months (Mean = 20.15)	ANOVA (IV: early talker (≥ 90th percentile), middle range talker, late talker (≤ 10th percentile) at 20 months[17])		Effect size not reported significant main effect (p < .0001), post hoc Tukey test revealed that early talkers (mean = 230.96) already significantly differed from middle range talkers

[14] Oxford CDI is a UK adaption of the CDI:WG which is not standardized for the UK. It does not include a gesture subscale.
[15] Results remained similar after controlling for age
[16] Mean length of utterance in morphemes
[17] Gender-specific norms were used to establish language status

						(mean = 103. 73) and late talkers (mean = 79.70) in terms of comprehension scores at time 1; middle range talkers and late talkers did not differ significantly from each other
	CDI:WG (Words understood and also produced)	Range between 10 and 16 months (Mean = 13.45)	CDI:WS (Production)	Range between 16 and 25 months (Mean = 20.15)	ANOVA (IV: early talker (≥ 90th percentile), middle range talker, late talker (≤ 10th percentile) at 20 months[18])	Effect size not reported significant main effect ($p < .0001$), post hoc Tukey test revealed that early talkers (mean = 41.41) at time 2 had significantly higher scores in terms of the percentage of words understood and also produced at time 1 than middle range (mean = 16.47) and they scored significantly higher than late talkers (mean = 7.69)
44	CDI:WG (Comprehension)	Range between 10 and 16 months (Mean = 13.45)	CDI:WS (Production)	Range between 16 and 25 months (Mean = 20.15)	T-test (IV: early talkers who remained early talkers versus early talkers who regressed towards the mean at 20 months)	Eta squared = .25 (large) Children who remained early talkers (mean = 230.13) had significantly higher comprehension scores at time 1 compared to those who regressed towards the mean at time 2 (mean = 138.32)
	CDI:WG (Words understood and also produced)	Range between 10 and 16	CDI:WS (Production)	Range between 16 and	T-test (IV: early talkers who remained early talkers versus early	Eta squared = .47 (large) Early talkers who remained advanced (mean

[18] Gender-specific norms were used to establish language status

		months (Mean = 13.45)		25 months (Mean = 20.15)	talkers who regressed towards the mean at 20 months)		= 51.93) had significantly higher scores at time 1 compared to those who regressed towards the mean at time 2 mean = 30.18).
	CDI:WG	Range between 10 and 16 months (Mean = 13.45)	CDI:WS (late-talking status)	Range between 16 and 25 months (Mean = 20.15)	Stepwise logistic regression	Predictor variables: Time 1 age, Time 2 age, SES, Time 1 comprehension, production, words understood and produced and gestures	Comprehension did not contribute any unique variance when entered as last step Percent of words understood and also produced: $R2$ = .34 (small) Entire model $R2$ = .15 (moderate) (gestures contributed highest unique variance $R2$ = .06 (small))
	CDI:WG	Range between 10 and 16 months (Mean = 13.45)	CDI:WS (early-talking status)	Range between 16 and 25 months (Mean = 20.15)	Stepwise logistic regression	Predictor variables: Time 1 age, Time 2 age, SES, Time 1 comprehension, production, words understood and produced and gestures	Comprehension: $R2$ = .02 (very small) Words understood and produced did not contribute unique variance Entire model $R2$ = .40 (large) of which only comprehension and production ($R2$ = .02) contributed unique variance

3.6.3. Is infant vocabulary comprehension associated with later total language scores?

Most of the following studies (see Table 10) made use of low-income samples (Fish & Pinkerman, 2003; Hamadani et al., 2010; Kreisman, 2012), hence the results may not be generalizable to the general population. The previous section studied the relationship between infant comprehension and later production and found mostly moderate relationships between comprehension at 12-14 months and production at 24 months. When studying the relationship between comprehension and later overall language weak correlations were found between these ages (Kreisman, 2012), which may be due to the instrument used (Bayley language factor) as reported previously stronger associations were usually found between CDI measures.

At kindergarten age, the results of the correlations varied somewhat in strengths, as Fish and Pinkerman (2003) found weak and non-significant correlations with language at 4 and 5 years but Hamadani (2010) described moderate relationships between comprehension at 18 months and language at 5 years. This finding by Hamadani (2010) may be due to the measure used at follow-up which was an IQ instrument in comparison to the receptive and expressive language tests used in other studies. Furthermore, the children in the study by Hamadani (2010) "were extremely disadvantaged, and many were at least moderately stunted" (p. S203), this may have had a biological foundation and as found for children at family risk for dyslexia may have lead to higher stability in scores over time (Unhjem et al., 2015).

Whilst infant comprehension at 14 months could explain some small variance (2%) in later language at 5 years, receptive language contributed 8% at 2;6 years and 14% at 3;6 years (highest unique variance) to later language as part of the Finnish Jyväskylä Longitudinal Study of Dyslexia which followed-up around 200 children, half of which were at family risk for dyslexia (H. Lyytinen et al., 2001). Furthermore, environmental influences such as parents' education (8%) were also important but behaviours such as shared book reading (3%) only contributed small variances to later language ability. This showed that comprehension was important for later language but comprehension skills during toddlerhood were better predictors than during infancy for predicting language scores at kindergarten age.

Table 10. *Relationship between comprehension (using the CDI:WG measure) and later total language scores*

Authors (date)	N	Predictor measures	Age at predictor	Outcome measures	Age at outcome	Analyses	Covariates	Effect size
Kreisman (2012)	1458	CDI:WG	14 months	Bayley language factor (overall language)	24 months	Correlation		r = .19 (small)
Fish and Pinkerman (2003)	98	CDI:WG	15 months	PLS-3 (overall)	4 and 5;4 years	Correlations		4 years: r = .19 (weak) 5;4 years: r = .14 (weak) Neither correlation was statistically significant
Hamadani et al. (2010)	1429	CDI:WG	18 months	verbal IQ (Wechsler Preschool and Primary Scale of Intelligence, WPPSI)	5 years	Correlation	Controlling for age at 18 months	r = .37 (moderate)
Lyytinen et al. (2001)	200 (children with (N= 107) and without familial risk for dyslexia)	CDI:WG	14 months	Overall language skills (a composite score of expressive and receptive language consisting of Peabody Picture Vocabulary Test, Inflectional Morphology Test, Wechsler Preschool and Primary Scale of Intelligence-Revised: Vocabulary and Comprehension)	5 years	Regression	parental education, child's symbolic play, mother's symbolic language, receptive and expressive language between 14 months and 3;6 years, risk or control status	R2 = .02 (small) (highest unique variance was receptive language at 3;6 years (R2 = .14))

3.6.4. Summary

Comprehension was significantly correlated with later language scores. The strongest correlations were found between early and later comprehension scores if both tests were administered during infancy (effect sizes were moderate to large). The relationships were generally stronger when the follow-up measure was also a CDI instrument rather than another standardised language test. As found for comprehension and Phrases Understood, the associations were also significant with later expressive language, but stronger relationships were found between comprehension and production than with grammar. The strength of association slowly decreased over time if infant comprehension was correlated with language scores at 24 months and later. To successfully predict scores at kindergarten age, comprehension scores during toddlerhood rather than infancy should be used (see Cochet & Byrne, 2016; H. Lyytinen et al., 2001).

3.7. Is early infant vocabulary associated with later language and literacy?

Even though at the age of 5 years, children with an initial delay (late comprehenders or late producers) had caught up with their peers on most standardised language tests, delayed children scored significantly lower in terms of phonological working memory (see Table 11) and their skills were comparable with those children with SLI (Ellis Weismer, 2007). Delays were also visible on the Minnesota CDI for Letters and Numbers. Both delayed groups scored significantly lower than typical children but only late comprehenders continued to score below the typical range on the expressive and receptive subscales of the Minnesota CDI. The authors concluded that language delayed children continued to have weaknesses in language which were likely to last into adolescence (Rescorla, 2009; Tomblin, Zhang, Buckwalter, & O'Brien, 2003) and that low comprehenders (i.e. children who had comprehension and production delays at 16 months) were likely to be affected more strongly than low producers (Thal et al., 2013).

Other research also found that the relationships between vocabulary using the Oxford CDI (latent variable: comprehension and production) at around 16-24 months and different language and literacy variables five years later (i.e. age ranged between 4 and 9 years) were significant and mostly moderate in strengths when examining a group of 300 British children from above average SES backgrounds (Duff, Reen et al., 2015). When two predictors were included (i.e. infant vocabulary and familial risk) in the model, they accounted for 16% in variance in later vocabulary, 6% in phonological awareness, 21% in reading accuracy, 30% in reading comprehension. Even though this could be interpreted as a substantial variance explained by the predictor variables, the authors concluded that the contributions of infant vocabulary and familial risk status were insufficient when trying to make predictions for individuals.

Overall, infant vocabulary (infant data of comprehension and production) helped to significantly predict later language and literacy scores. Whilst early vocabulary could explain some variance in later vocabulary, it was even better at predicting later literacy ability. This was because most late producers and comprehenders had caught up with their peers in terms of expressive and receptive language by kindergarten age (Thal et al., 2013), but they continued to lag behind their peers in the newly developing skill of literacy.

Table 11. *Relationship between vocabulary (using the CDI:WG measure) and later language and literacy*

Authors (date)	N	Predictor measures	Age at predictor	Outcome measures	Age at outcome	Analyses	Covariates	Effect size
Duff et al. (2015)	300	Oxford CDI (latent variable: comprehension and production)	Range between 16 – 24 months	Receptive and Expressive One Word Picture Vocabulary Test[19] (latent variable: receptive and expressive vocabulary)	6;09 years (Range between 4 and 9 years)	Structural Equation Model (SEM): Correlation		r = .40 (moderate)
				Elision subtest of the Comprehensive Test of Phonological Processing[20] (phonological awareness)		SEM: Correlation		r = .21 (weak)
				Diagnostic Test of Word Reading Processes[21] (reading accuracy)		SEM: Correlation		r = .33 (moderate)
				York Assessment of Reading Comprehension[22] (reading comprehension)		SEM: Correlation		r = .43 (moderate)

[19] Brownell et al. (2000a; 2000b)
[20] Wagner, Torgesen and Rashotte (1999)
[21] Forum for Research into Language and Literacy (2012)
[22] Snowling et al. (2009)

				Receptive and expressive vocabulary / phonological awareness/ reading accuracy)/ reading comprehension		SEM	predictor variables: infant vocabulary and familial risk	$R2 = .16$ (moderate) $R2 = .06$ (weak) $R2 = .21$ (moderate) $R2 = .30$ (large)
Thal et al. (2013)[23]	64 at 4 years (44 typical children, 20 late producers/comprehenders) 65 at 5 years (44 typical children, 21 late producers/comprehenders)	CDI:WG	16 months	non-word repetition test (NRT) which tests phonological working memory	4 and 5 years	ANOVA (IV: delayed group (late comprehenders, late producers) and typical group; DV: overall percent of phonemes correct (TPPC) and 4 subscales (phonemes correct for 1-4 syllable level)		Effect sizes not reported 4 years: delayed group performed significantly worse on two out of the four scales (3-and 4-syllable level) and TPPC 5 years: delayed group scored significantly lower on the 4- syllable level and the TPPC compared to the typical group
	90 (individual group sizes not specified)			Minnesota CDI for Letters and Numbers	5 years	ANOVA (IV: late comprehenders, late producers typical group)		Effect size not reported late comprehenders and late producers did not differ between themselves but both groups scored significantly lower than the typical group

[23] At 4 and 5 years, the mean values of the delayed groups were within the normal range for both tests (NRT and Minnesota CDI).

3.8. CDI:WG - Production

3.8.1. Is early vocabulary production associated with later production (and grammar)?

Infant production was correlated with later production and grammar scores, see Table 12. The relationships between infant production and later production and grammar scores up to 24 months were moderate and large in strengths, with few exceptions (Unhjem et al., 2015). This relationship could also be seen when studying differences in ability groups (late, middle and early talkers or late producers, late comprehenders and typical children) with language scores at earlier or later time points. For example, Thal et al. (1997) found significant differences between early, middle and late talkers several months earlier in the expected direction (early talkers produced most words followed by middle range talkers who produced significantly more words than late talkers). They also reported significant differences within groups as children who remained early talkers at 20 months produced significantly more words at 13 months than those early talkers who regressed towards the mean. The stability of production scores was further supported as correlations did not differ depending on reporter type (e.g. mother, father, see Bornstein et al., 2006) and the association between time points remained significant even after controlling for gender, SES, age, birth order, ethnicity and comprehension (e.g., Feldman et al., 2000; P. Lyytinen et al., 1999); however only from 12 months due to floor effects in production scores for younger children (Fenson et al., 1994). Furthermore, early-talking status could be explained in part by previous production and comprehension scores; however, late-talking status could not be explained by earlier production skills, instead gestures and the *percentage of words understood and also produced* added unique variance to late talking status (Thal et al., 1997).

It should be noted that production at follow-up was usually tested using the CDI:WS production subscale. Only the study by Unhjem et al. (2015) used composite scores for production combining scores of the CDI:WS with production scores of the CDI:WG or the expressive Bayley scores. This may be the reason for not finding significant correlations as the CDI:WS was the only instrument out of these three which tested age-appropriate word production skills at follow-up. If ceiling effects were reached due to an age-inappropriate instrument (CDI:WG) or general language production (expressive Bayley scores) was tested, it was likely that correlations were less strong. However, in contrast to the control children, for children at-risk for dyslexia all associations were large between 12 and 24 months regardless of instrument used at follow-up. Furthermore, expressive Bayley scores could be better predicted from previous CDI scores for at-risk children compared to control children. Again, this showed that language was more stable for children with familial risk for dyslexia compared to typically developing children.

When studying the relationship between infant production scores and production during the third year (children aged around 2;6 years), it was expected that the strengths of association decreased with an increase of time between first testing and follow-up testing. However, Cochet & Byrne (2016) reported large effect sizes, see above for possible reasons. Thal et al. (2013) studied children between 16 months (CDI:WG) and 28 months (CDI:WS) and found moderate relationships between CDI:WG production and CDI:WS production as well as grammar. They also added that language delay at 28 months could be predicted significantly but weakly from production scores at 16 months after controlling for a range of factors (e.g. ear infections, family history of speech and language disorders and learning disabilities, parental education and income). However, Guiberson (2008) did not find any significant associations with CDI:WS production scores or with observed words in mother-child interactions. The reason for this could be that the study used the INV questionnaires which was standardised for Mexican children but for their study Guiberson (2008) looked at Mexican immigrants to the US with mostly monolingual Spanish speaking parents from low SES backgrounds. It is not clear if the same CDI adaptations were suitable for the use in the US due to a different cultural environment compared to Mexico. For example, it is possible that the new environment caused parents to use different objects, toys or animals compared to their home environment which in turn may have had an effect on the frequency of word use. Thus, the standardised INV questionnaires may not have been appropriate for Mexican immigrants in the US and may not have given a true representation of the children's language ability.

Looking at the developmental trajectory between 16 and 36 months of late producers, late comprehenders and typically developing children, all groups developed their knowledge of vocabulary and grammar over time but at different rates (Thal et al., 2013). During the third year, late producers had scores in the normal range (35th to 55th mean percentiles) and had caught up with typically developing children in terms of grammar. While late comprehenders (children with below cut-off comprehension AND production at 16 months) also scored within the typical range, they still lagged behind significantly in comparison to the other two groups.

The relationship between infant production scores and production at 36 months varied depending on the study between significant and weak (using the ETS test, see Rose et al., 2009) or non-significant associations (using the Renfrew lexical skills, see Korpilahti et al., 2016). The results may have depended on the production test used at follow-up as Rose et al. (2009) found that when predicting ETS scores at 36 months, the contributions of production at 12 months remained significant even after controlling for prematurity and after adding information processing skills as a last step to the model. Interestingly, Korpilahti et al. (2016) reported that while production at 13 months could not explain any significant variance in later expressive language, production at 24 months was the only significant predictor.

As the initially delayed children had caught up with their peers at around 5 years in terms of standardised expressive language tests (Thal et al., 2013), some significant and weak associations could still be found between infant production and later production scores at 4 years but not anymore at 5 years (Fish & Pinkerman, 2003). This could also be seen from the study by Duff et al. (2015) who found that British English average talkers and late talkers (established at 18-19 months) from middle to upper middle class backgrounds did not significantly differ in terms of language and literacy scores at around 4-9 years. The results showed that most of the average talkers stayed in the average to above average category whilst most of the late talkers moved up into to the average to high-average category. However, the prediction was very good when SLI status at 4 years was used to predict scores at around 8-9 years of age. Out of the late talkers and the average talkers, a similar number of children moved into the SLI category. Even though Duff et al. (2015) did not find any significant differences between late and average talkers at 8 years, it is possible that the incorporation of other CDI measures (i.e. gestures, comprehension) as predictors or mediating factors could have given a better insight. The authors also recognised that they did not incorporate any grammar scales at follow-up although which could have added another dimension to the results. It would have been interesting to find out if the percentage of children affected by SLI was different for children who were precocious at 18 months and if they were less likely to fall behind.

In summary, moderate to large relationships existed between early production (from around 12 months but not earlier) and later production or grammar during infancy. This associations were weak if follow-up testing occured at 3 years or later. This was because most late talkers had caught up with their typically developing peers in terms of grammar and vocabulary while they still lagged behind in other language domains at 4 and 5 years (e.g., phonological working memory, see Thal et al., 2013). Furthermore, late talkers as well as typically developing children could develop SLI during kindergarten age. Due to the high recovery rate of late talkers (50 -75%, see Bavin & Bretherton, 2013; Paul, 2000; Rescorla et al., 2000; Roos & Ellis Weismer, 2008), SLI status may be a better predictor for language ability during middle childhood (Duff et al., 2015).

Table 12. *Relationship between production (using the CDI:WG measure) and later production*

Authors (date)	N	Predictor measures	Age at predictor	Outcome measures	Age at outcome	Analyses	Covariates	Effect size
Thal et al. (2013)[24]	1107 (863 typically developing children, 154 late producers with typical comprehension, 90 late comprehenders (≤ 10th percentile for comprehension and production)	CDI:WG	10 and 13 months	CDI:WG	16 months	Descriptive data (IV: typically developing, late producers (≤ 10th percentile), late comprehenders (≤ 10th percentile) at 16 months)		Effect size not reported (descriptively) Late producers identified at 16 months already showed smaller productive vocabularies at 10 and 13 months compared to typically developing children but a bigger vocabulary size compared to late comprehenders
		CDI:WG	16 months	CDI:WS	28 months	Three-way ANOVA with same IVs as above separate for 3 different DVs (vocabulary, sentence complexity, mean of the 3 longest utterances (M3L))		For all three ANOVAS, late producers had significantly lower scores than typical children but both groups scored significantly higher than late comprehenders: amount of variance accounted for was 23% for vocabulary size, 20% for sentence complexity and 10% for M3L
		CDI:WG	16 months	CDI:WS	28 months	Correlations		Production: r = .45 (moderate) Sentence complexity: r = .50, (moderate to strong)
	391 (262 typically developing children, 86 late	CDI:WG	16 months	CDI:WS (word combinations)	20, 24, 28 and 36 months	Three-way ANOVA		No effect sizes reported late producers scored significantly lower than typical children but did not differ

[24] Children from all groups scored within the normal range at 28 months and at 4 and 5 years

producers, 43 late comprehenders)							from late comprehenders at 20 and 24 months, the three groups did not differ from each other at 28 and 36 months
			CDI:WS (complex sentences)				At 20 months: late producers did not significantly differ from the other groups (floor effects) At 24 and 28 months: late producers scored significantly lower than typical children and significantly higher than late comprehenders At 36 months: late producers and typical children did not differ and both groups scored significantly higher than late comprehenders
			CDI:WS (M3L)				At 20, 24 months: late producers scored significantly lower than typical children but did not differ from late comprehenders At 28 months: late producers scored significantly lower than typical children and significantly higher than late comprehenders At 36 months: late producers and typical children did not differ and both groups scored significantly higher than late comprehenders
	CDI:WG	16 months	CDI:WS (language delay ≤ 10th percentile for vocabulary	28 months	regression	Predictor variables: child's history of ear infections by 16 months, family history of speech	Unique contribution of production had the Somer's D = .22 (weak)

				production AND/OR sentence complexity score)			and language disorders, family history of learning disabilities, maternal and paternal education, family income, CDI:WG at 16 months comprehension, production, gesture	
	90 (group sizes not specified)			Expressive Vocabulary Test (EVT)[25]	4 and 5 years			4 years: Late producers scored significantly lower than typical children 5 years: late producers did not differ from typical children and late comprehenders
	90 (group sizes not specified)			Expressive subscale of the Clinical Evaluation of Language Fundamentals Preschool (CELF-P)[26]	4 and 5 years			At 4 years: Late producers scored significantly lower compared to typical children but did not differ from late comprehenders At 5 years: Late producers did not differ from late comprehenders and typical children
Fenson et al. (1994)	62	CDI:WG	9 months (range 8.03 - 10.97 months)	CDI:WG	16 months (range 14.93 - 18.63 months)	Correlation		r = .32 (moderate)

[25] Williams (1997)
[26] Wiig, Secord & Semel (1992)

						Stepwise regression	age, gender, birth order, SES	Production: R^2 = .08 (weak) (The model did not reach significance as a whole)
		CDI:WG	13 months (range 10 - 16 months)	CDI:WS	20 months (range 16 - 25 months)	Correlation		r = .69 (large)
						Stepwise regression	age, gender, birth order, SES	R^2 = .25 (large)
Fenson et al. (2007)	62	CDI:WG	9 months (range 8 - 10 months)	CDI:WG	16 months (range 14 – 17 months)	Correlation		r = .38 (moderate)
	217	CDI:WG	Range 10 - 16 months	CDI:WS	Range 16 – 25 months	Correlation		r = .69 (large)
Sachse et al. (2007)	149	ELFRA-1	12 months	ELFRA-2	24 months	Correlation		rs = .41 (moderate)
Feldman et al. (2000)	2156	CDI:WG	12 months	CDI:WS	24 months	Correlations		Production: r = .39 (moderate) Grammar: mostly moderate associations
				CDI:WS (production)	24 months	Multiple regression	child’s age at time 2, child’s sex, maternal education, health insurance status, race	R^2 = .13 (moderate)

Bavin et al. (2008)	1467	CDI:WG	12 months	CDI:WS	24 months	Correlations	Production: r = .41 (moderate)
Bornstein et al. (2006)	29	CDI:WG	13 months	CDI:WS	20 months	Correlations	All associations between production scores were moderate when using different reporter types (mother, father etc.)
Kreisman (2012)	1458	CDI:WG	14 months	CDI:WS	24 months	Correlations	Production: r = .38 (moderate) Sentence complexity: r = .32 (moderate)
Hamadani et al. (2010)	1429	CDI:WG	12 months	CDI:WG	18 months	Correlation	Production: r = .50 (moderate to large)
Guiberson (2008)	19	CDI:WG	Range between 14 -16 months	Observed words in mother-child interactions	Range between 30 – 32 months	Regression	R2 = .02 (weak) Relationship was not statistically significant
		CDI:WG	Range between 14 -16 months	CDI:WS	Range between 30 – 32 months	Regression	R2 = .07 (weak) Relationship was not statistically significant
Unhjem et al. (2015)[27]	32 (at familial risk for dyslexia)	CDI:WG (at 18 months: composite production	12, 15, 18 months	CDI:WG at 15 months; composite CDI:WG + CDI:WS at 18 months; composite CDI:WS + Bayley at 24 months	15, 18, 24 months	Predictive correlations	All correlations between production scores were strong

[27] Fisher's z-transformed correlations coefficients showed that the correlation coefficients between the at risk and control group were not significantly different from each other.

		CDI:WG + CDI:WS)						
	21 (control children)	CDI:WG (at 18 months: composite production CDI:WG + CDI:WS)	12, 15, 18 months	CDI:WG at 15 months; composite CDI:WG + CDI:WS at 18 months; composite CDI:WS + Bayley at 24 months	15, 18, 24 months	Predictive correlations		Associations between production scores were mostly strong except for some non-significant associations (between 12 and 18 months, between 15 and 24 months)
	32 (at familial risk for dyslexia)			Bayley production	24 months	Hierarchical regression	step 1 CDI comprehension at 12 months, CDI production at 12 months; step 2 CDI comprehension 15 months, CDI production at 15 months; step 3 CDI comprehension at 18 months, CDI composite production at 18 months	Composite at 18 months R^2 = .30 (large) Production at 12 and 15 months were not significant predictors
	21 (control children)							Production at 12 months R^2 = .34 (large)(only significant predictor in model)
Lyytinen et al. (1996)	94	CDI:WG	14 months	CDI:WS	18 months	Correlations		Production: r = .76 (large) Use of suffixes: r = .53 (large)

				Expressive RDLS	18 months	Correlation		Production: r = .60 (large)
Lyytinen et al. (1999)	171	CDI:WG	14 months	CDI:WS	18 months	Correlations		Production: r = .73 (large)
		CDI:WG (production)	14 months	CDI:WS production CDI:WS Maximum sentence length Bayley expressive score	24 months	Multiple regression	At 14 months: maternal education, gender, CDI comprehension	R2 = .08 (weak) R2 = .10 (weak) R2 = .03 (weak)
		CDI:WS (production)	18 months	CDI:WS production CDI:WS Maximum sentence length Bayley expressive score	24 months	Hierarchical multiple regression	At 18 months: maternal education, gender, RDLS comprehension	R2 = .31 (large) R2 = .24 (large) R2 = .07 (weak)
Korpilahti et al. (2016)	226	CDI:WG	13 months	CDI:WS	24 months	Correlation		Production: r = .33 (moderate)
				Renfrew lexical skills	36 months	Correlation		r = .12 (weak and not statistically significant)
				Renfrew lexical skills	36 months	Multiple linear regression	CDI:WG production, comprehension at 13 months and CDI:WS production at 24 months	Production at 13 months did not contribute significant variance to the model (Entire model R2 = .15 (moderate) of which only production (CDI:WS) at 24 months was significant)

Rose et al. (2009)	182 (pre-term = 56), full-term = 126)	CDI:WG	12 months	Educational Testing Service (ETS) test of verbal fluency[28]	36 months	Correlation	controlling for birth status (preterm vs full-term birth)	r = .29 (weak)
						Hierarchical regression	Step 1 birth status, Step 2 infant language at 12 months (comprehension, production, Communicative Gestures)	Infant language (significant for comprehension and production) R2 = .12 After adding information processing as step 3 infant language remained significant (only production remained significant)
Fish and Pinkerman (2003)	98	CDI:WG	15 months	PLS-3	4 years 5;4 years	Correlations		4 years: r = .25 (weak) 5;4 years: r = .15 (weak and not statistically significant)
Cochet and Byrne (2016)	8	Oxford CDI	Range between 11 – 33 months	Oxford CDI	Range between 21 – 41 months	Correlations		Production: r = .87 (strong)
Duff, Nation, Plunkett, and Bishop (2015)[29]	60 (average versus (N= 30), late-talkers (N=30))	Oxford CDI	18 – 19 months (late-talking: 10 words or fewer equals 25th	Receptive and expressive Vocabulary: One Word Picture Vocabulary Test[31] Nonword repetition:	Range between 4 - 9 years (groups were age and	Matched paired t-tests		Mean language scores of late talkers were generally lower than for average talkers at follow up but the difference was not significant for most tests: Receptive vocabulary: d = .33 (small and not significant) Expressive vocabulary:

[28] Singer, Corley, Guiffrida and Plomin (1984)
[29] At follow-up all groups (groups classified at time 1: average talkers, mild and severe late talkers) had mean scores within or above the typical range
[31] Brownell (2000a; 2000b)

		centile of the Oxford CDI[30]; average talking: 31st to 73rd centile)	Children's Test of Nonword Repetition[32]; Recalling sentences: Clinical Evaluation of Language Fundamentals (CELF-III UK)[33]; Phonological elision: Comprehensive Test of Phonological Processing[34]; Reading accuracy: Diagnostic Test of Word Reading Processes[35], Reading comprehension: York Assessment of Reading Comprehension[36]	gender matched)		d = .58 (moderate and not significant after controlling for multiple comparisons) Nonword repetition: d = .33 (small and not significant) Recalling sentences: d = .27 (small and not significant) Phonological elision: d = -.14 (very small and not significant) Reading accuracy: d = .17 (very small and not significant) Reading comprehension: d = .25 (small and not significant)
	50	18 – 19 months	See above	Range between 4 - 9 years	Matched paired t-tests	Mean language scores of late talkers were generally lower than for average talkers at follow up but the difference was not significant for most tests:

[30] Hamilton, Plunkett & Schafer (2000)
[32] Gathercole & Baddeley (1996)
[33] Semel, Wiig & Secord (2000)
[34] Wagner, Torgesen & Rashotte (1999)
[35] Forum for Research into Language and Literacy (2012)
[36] Snowling et al. (2009)

(average versus (N= 30), late-talkers (N=20))		(late-talking: 6 words or fewer equals 16^{th} centile of the Oxford		(groups were age and gender matched)		Receptive vocabulary: d = .50 (moderate and not significant) Expressive vocabulary: d = .82 (large and not significant after controlling for multiple comparisons) Nonword repetition: d = .57 (moderate and not significant) Recalling sentences: d = .58 (moderate and not significant) Phonological elision: d = -.22 (small and not significant) Reading accuracy: d = .27 (small and not significant) Reading comprehension: d = .46 (small and not significant)
18 (average versus (N= 12), SLI (N=6))	SLI diagnosis (9 late talkers and 9 average talkers were tested at 4 years, of which 3 of each group	4 years	See above	Range 8;01-9;04 years	Independent samples t-tests	Mean scores were lower for the SLI group on all tests but within typical range: Receptive vocabulary: d = 1.04 (large) Expressive vocabulary: d = 2.30 (large)

		were diagnosed with SLI and the others were typical)					Nonword repetition: d = 1.95 (large) Recalling sentences: d = 2.24 (large) Phonological elision: d = .92 (large) Reading accuracy: d = 2.19 (large) Reading comprehension: d = 1.71 (large)
Thal et al. (1997)	217	CDI:WG (Production)	Range between 10 and 16 months (Mean = 13.45)	CDI:WS (Production)	Range between 16 and 25 months (Mean = 20.15)	ANOVA (IV: early talker (≥ 90th percentile), middle range talker, late talker (≤ 10th percentile) at 20 months[37])	Effect size not reported significant main effect ($p < .0001$), post hoc Tukey test revealed that all groups differed significantly from each other. Early talkers (mean = 102.65) produced most words at time 1 followed by middle range talkers (mean = 17.72) and late talkers (mean = 4.13).
	44	CDI:WG (Production)	Range between 10 and 16 months (Mean = 13.45)	CDI:WS (Production)	Range between 16 and 25 months	T-test (IV: early talkers who remained early talkers versus early talkers who regressed towards the mean at 20 months)	Eta squared = .46 (large) Children who remained early talkers at time 2 (mean = 131.00) produced significantly more words at time 1 than those early talkers who

[37] Gender-specific norms were used to establish language status

				(Mean = 20.15)			regressed towards the mean at follow-up (mean = 40.07)
	CDI:WG	Range between 10 and 16 months (Mean = 13.45)	CDI:WS (late-talking status)	Range between 16 and 25 months (Mean = 20.15)	Stepwise logistic regression	Predictor variables: Time 1 age, Time 2 age, SES, Time 1 comprehension, production, words understood and produced and gestures	Production did not contribute any unique variance when entered as last step Entire model R2 = .15 (moderate) (only gestures and words understood and produced added unique variance to the model)
	CDI:WG	Range between 10 and 16 months (Mean = 13.45)	CDI:WS (early-talking status)	Range between 16 and 25 months (Mean = 20.15)	Stepwise logistic regression	Predictor variables: Time 1 age, Time 2 age, SES, Time 1 comprehension, production, words understood and produced and gestures	Production: R2 = .02 (small) Entire model R2 = .40 (large) of which only production and comprehension (R2 = .02) contributed unique variance

3.8.2. Is early vocabulary production associated with later comprehension?

The relationships between infant production and later comprehension was generally less strong than with later production scores. The associations were of weak to moderate strengths if comprehension was assessed at around at 18 months (P. Lyytinen et al., 1996; Unhjem et al., 2015), see Table 13. Whilst the associations were not statistically significantly different between the control children and children at familial risk for dyslexia, the associations were strong for at-risk children (Unhjem et al., 2015).

The relationships between infant production and comprehension scores at 36 months remained weak to moderate (Korpilahti et al., 2016; Rose et al., 2009) and significant, except for the study by Cochet & Byrne (2016) who found non-significant relationships but they used a very small sample (N=8) and employed the Oxford CDI for measuring comprehension at an age when children were too old for the questionnaire (21-41 months). Furthermore, the relationship remained significant after controlling for prematurity; however, after including infant information processing as a last step only comprehension and not production remained significant as a predictor of infant language. This was because infant comprehension and other skills such as infant cognitive ability (i.e. information processing) were more strongly correlated with later comprehension than infant production (Rose et al., 2009). As reported previously later language scores could generally be best predicted from language data of older rather than younger children, this was supported by the regression analysis of Korpilahti et al. (2016).

Most of the studies reported so far used production scores from children aged 12 – 14 months which could be the reason for the somewhat low correlations with later comprehension scores. This was because children who were later classified as late comprehenders or late producers both showed floor effects in early production when tested at 10 and 13 months (Thal et al., 2013) but whilst late producers did not have comprehension deficits, late comprehenders remained delayed in comprehension and production. Therefore, lower associations were expected as the groups were naturally developing in different directions.

At around 4 to 5 years, late-talking children had caught up with their typically developing peers in terms of test scores on standardised comprehension tasks. Even though late producers scored within the typical range at this age, they still had significantly lower scores than typical children on some tests (Thal et al., 2013). This could explain why production at 15 months (when language was more developed than for the correlations reported above between 12 - 14 months) was a significant predictor of language comprehension scores at 5 years (Fish & Pinkerman, 2003).

Table 13. *Relationship between production (using the CDI:WG measure) and later comprehension*

Authors (date)	N	Predictor measures	Age at predictor	Outcome measures	Age at outcome	Analyses	Covariates	Effect size
Lyytinen et al. (1996)	94	CDI:WG	14 months	RDLS verbal comprehension	18 months	Correlation		r = .20 (weak)
Korpilahti et al. (2016)	226	CDI:WG	13 months	RDLS III	36 months	Correlation		r = .25 (weak)
		CDI:WG (comprehe nsion)	13 months	RDLS III	36 months	Multiple linear regression	predictor variables: CDI:WG Production, Comprehension at 13 months, CDI:WS Production at 24 months	Comprehension at 13 months did not contribute significant variance (only Production at 24 months was significant R2 = .19)
Unhjem et al. (2015)[38]	32 (at familial risk for dyslexia)	CDI:WG	12, 15 months	CDI:WG	15, 18 months	Longterm correlations		All correlations were significant and strong
	21 (control children)	CDI:WG	12, 15 months	CDI:WG	15, 18 months	Longterm correlations		The only significant correlation was between 12 and 18 months and moderate in strength (r = .45)
Rose et al. (2009)	182 (pre-term = 56), full-term = 126)	CDI:WG	12 months	PPVT	36 months	Correlation	controlling for birth status (preterm vs full-term birth)	r = .40 (moderate)
						Hierarchical regression	predictor variables: Step 1 birth status,	Infant language R2 =.26 (large, comprehension and production were

[38] Fisher's z-transformed correlations coefficients showed that the correlations coefficients between the at risk and control group were not significantly different from each other.

							Step 2 infant language at 12 months (comprehension, production, Communicative Gestures)	significant) after adding infant information processing as step 3 comprehension but not production remained significant
Cochet and Byrne (2016)	8	Oxford CDI	Range between 11 – 33 months	Oxford CDI	Range between 21 – 41 months	Correlation		rs = .54 (moderate but not statistically significant)
Fish and Pinkerman (2003)	98	CDI:WG	15 months	PLS-3	4 years	Correlation		r = .14 (small and not statistically significant)
				PLS-3	5;4 years	Correlation		r = .28 (weak)
				PLS-3	5;4 years	Hierarchical regression	predictor variables: step 1 articulation problems at 5;4 years, uncooperative behaviour at 5;4 years, child's sex, step 2 attachment security at 15 months, step 3 child's number of books at 4 years, maternal prediction of child's school success, step 4 child's number of words produced at 15 months, child's lack of initiative at 4 years, step 5 child's facilitation at 9 months, maternal over-control at 4 years	Step 4: $R2$ = .13 (moderate) both predictors remained significant after all predictors had been added

Thal et al. (2013)[39]	90 (individual group sizes not specified)	CDI:WG	16 months	Peabody Picture Vocabulary Test - Revised	4 and 5 years	Three-way ANOVA (IV: typically developing, late producers (≤ 10^{th} percentile), late comprehenders (≤ 10^{th} percentile) at 16 months)		Effect sizes not reported 4 years: groups did not differ significantly 5 years: late producers had significantly lower scores compared to typical children but did not differ from late comprehenders
				CELF-P Receptive	4 and 5 years			Effect sizes not reported 4 years: late producers had significantly lower scores compared to typical children but did not differ from late comprehenders 5 years: groups did not differ significantly
				Minnesota CDI	5 years			Effect sizes not reported Late producers did not significantly differ from typical children

[39] The means of late producers were within the normal range at 4 and 5 years on all three tests described here.

3.8.3. Is early vocabulary production associated with later total language scores?

Infant production scores correlated significantly but weakly with language scores up to kindergarten age, see Table 14. Most studies reported here presented results from low-income samples (Fish & Pinkerman, 2003; Hamadani et al., 2010; Kreisman, 2012), thus the findings could be difficult to generalise to the population as a whole. An exception was the study by Hsu and Iyer (2016) who made use of a representative American sample. They found that production at 15 months (but not comprehension or communicative gestures) significantly predicted risk for language impairment at 3 to 4 years of age. Children who were at risk for language impairment at 3 years produced significantly fewer communicative gestures, comprehension and production scores at 15 months; however, children at risk for LI at age 4;6 only showed significantly lower productive vocabularies at 15 months (Hsu & Iyer, 2016). A structural equation model confirmed that word production at 15 months predicted later risk for language impairment (composite of at-risk status at 3 and 4;6 years). Furthermore, the influence of early gestures on later risk for language impairment was mediated by production skills at 15 months. This was an interesting finding as late-talking at 20 months could be predicted in part from gestures scores at 13 months (see Table 12) in the study by Thal et al. (1997); once children got older production became more important for predicting later language status but as can be seen by Hsu and Iyer (2016) gestures still had an effect albeit an indirect effect. Thus, it seems that the age of testing and the language measure used at follow-up were important to make predictions about outcomes.

Table 14. *Relationship between production (using the CDI:WG measure) and later total language scores*

Authors (date)	N	Predictor measures	Age at predictor	Outcome measures	Age at outcome	Analyses	Covariates	Effect size
Kreisman (2012)	1458	CDI:WG	14 months	Bayley language factor	24 months	Correlation		r = .23 (small)
Hamadani et al. (2010)	1429	CDI:WG	18 months	Verbal IQ measure of the Wechsler Preschool and Primary Scale of Intelligence	5;3 years	Correlation		r = .32 (moderate)
Fish and Pinkerman (2003)	98	CDI:WG	15 months	PLS-3 total	4 years	Correlation		r = .22 (small and not significant)
					5;4 years	Correlation		r = .23 (small)
Hsu and Iyer (2016)	1064 (of which at 3 years 149 and at 4;6 years 226 at risk for LI)	CDI:WG	15 months	At risk group at 3 years: at or below 10th percentile on RDLS Verbal Comprehension and/or Expressive Language scale At risk group at 4;6 years: at or below 10th percentile on PLS-3 Auditory Comprehension and/or Expressive Communication scale	3 and 4;6 years	ANOVA		Eta squared not reported Children with risk of language Impairment at 3 or 4;6 years had significantly lower production scores at 15 months compared to typical children
				Composite risk of LI at 3 and 4;6 years	composite of at-risk status at 3 and 4;6 years	SEM	Communicative gestures Production Comprehension	Only Production at 15 months directly predicted later risk for language impairment (total model: R2 = .43 (large))

3.8.4. Summary

In general, early production scores were significantly correlated with later language. The strongest correlations (with moderate to large effect sizes) were reported between early production scores (at around 12 months) and later production or grammar scores (at around 24 months). In most cases, early production scores contributed the highest additional variance to later expressive language outcomes for early talkers or samples with mixed abilities but not for late-talkers. This meant that early production scores on their own might have not been sufficient to predict slow production development during the second year. Furthermore, in studies which predicted language ability at kindergarten age and middle childhood, production scores at 24 months and later (Duff et al., 2015; Korpilahti et al., 2016) were better predictors than production during infancy.

3.9. Individual-level analyses: Predicting outcomes for individual children

Most statistical analyses exemplified above make use of correlations, regressions or group mean comparison designs – these are group-level analyses to establish measures of association and central tendency across a wide spectrum of ability (Dollaghan, 2013). Even though these techniques help to provide evidence for distinction between categories, they are not sensitive enough to make clinical predictions for individuals (Dollaghan, 2013). For an index of classification accuracy for individual-level analyses 2 x 2 contingency tables are conventionally used (first column: disease is present; second column: disease is absent; upper row: test result was positive/exposed; lower row: test result was negative/unexposed) to describe the numbers of true positives (a), false negatives (b), false positives (c) and true negatives (d). Accuracy values can be reported in different ways as odds ratios ((a x d)/(c x b)), relative risk ratios ((a/(a+c))/(b/(b+d))), sensitivity (a/(a+b)), specificity (d/(c+d)), positive predictive value (a/(a+c)), negative predictive value (d/(b+d)), and positive (sensitivity/(1-specificity)) and negative likelihood ((1-sensitivity)/specificity) ratios (MEDCALC, 2017). Hence, the information from the contingency tables indicates how many of the delayed children were correctly classified at earlier stages (sensitivity) and how many of the children with typical development were already categorised as typical at earlier stages (specificity).

In order to predict classification accuracy the following studies used different types of analyses. The types of analyses used were discriminant analysis (Fish & Pinkerman, 2003; Thal et al., 1997), ROC-curves (Westerlund, Berglund, & Eriksson, 2006), logistic regression (Unhjem et al., 2015) or 2 x 2 contingency tables (Sachse, Saracino et al., 2007).

Discriminant analysis was used by Fish and Pinkerman (2003) to test if children with low SES could be correctly classified as below average language ability (total language score <85) at 4 years and 5;4 years of age. It should be said here that children within the normal range score between 85 and 115 and a score of 130 is only achieved by the best 2% of the population (98th percentile). The predictor variables were all significant (at 4 years: child's lack of initiative at 4 years, maternal over-control at 4 years, maternal facilitation at 9 months, child's words produced at 15 months, child's number of books at 4 years; at 5;4 years: child's number of books at 4 years, child's PLS-3 total score at 4 years, attachment security at 15 months, child's words produced at 15 months). As many of the predictor variables were collected at 4 years, it may not be surprising that more children (4 years: 74.4% correctly classified, 5;4 years: 76;8% correctly classified) could be correctly classified than in Thal et al.'s study, but it still means that around a quarter of children were misclassified.

Thal et al. (1997) utilized the US norm data and also made use of discriminant analysis in order to test how many children could be correctly classified as early or late talkers at 20 months on the CDI:WS. They used seven predictor variables including CDI:WG scores collected at 13 months (predictor variables: time 1 age, time 2 age, SES, comprehension, production, gestures, percentage of all words in receptive vocabulary which are also produced). The proportions of overall correct classification (i.e. no separate values for specificity and sensitivity) and separate false/true negatives and positives were reported. It was found that around two-thirds of children could be correctly identified using the predictor variables which means that many children were also misclassified. In the analysis of late talker status, the chi-square statistic was reliable (chi square = 34.86, $p < .00001$) indicating that within-group variability was small compared to between-group variability.

Another study by Thal et al. (2013) followed children from 10 months to 7 years in order to examine if language delay could be identified at 16 months or earlier. They found that 6.7% of children in their sample (N= 926) were diagnosed with a language disorder at some point between 4 and 7 years of age. Similar prevalence rates have been reported in other studies (e.g., Norbury et al., 2016; Tomblin et al., 1997). At 16 months children were classified into typically developing children (N = 365), late producers (N = 70) and late comprehenders (N = 28). The rates of language disorder identified between 4 and 7 years varied a lot depending on group status at 16 months, that is the rates were a lot higher in the late comprehender group (17.8%) compared to the late producer (11.4%) and the typically developing children (4.9%). There was no significant difference between the percentage of diagnoses for the two delayed groups but they differed significantly from the typical group.

A subsample (N = 470) of which 6.81% (N=32) had received a diagnosis of language impairment between 4 and 7 years were used for the following analysis. A stepwise logistic regression examined the strength of prediction for the language impairment outcome by using several communication variables at 16 and 28 months as predictors (CDI:WG production, comprehension and gestures, CDI:WS production and sentence complexity). The best model fit included the predictors: production at 16 months, production at 28 months and sentence complexity at 28 months. The predictive strength was moderate (Somer's D = .49). The classification accuracy was assessed using different cut-off points depending on prior probability of language disorders. Prior probability values were calculated from the continuous CDI variables. Low prior probability values achieved higher sensitivity levels and high probability values achieved higher specificity levels. Depending on which cut-off was used the values varied (from low to high probability levels) for sensitivity: 69% (i.e. 22 out of 32 children correctly identified with a delay) to 41% (i.e. 13 out of 32 children correctly identified with a delay) and specificity: 69% (i.e. around 302 out of 438 children correctly identified with no delay) to 95% (i.e. around 416 out of 438 children correctly identified with no delay). Predictive value positive (14 to 37%) and predictive value negative (97% to 95%) were also given. Whilst the specificity value can be reduced to 69%, the sensitivity value rises at the same time (69%) which means that a high number of children can be identified correctly with a language problem during toddlerhood (69%, 22 children) there is also a high percentage of children (31%, 136 children) who would be incorrectly identified. On the other hand if a higher probability value is used, the sensitivity would be lower (41%, 13 children), but only a small percentage of toddlers would be incorrectly identified (5%, 22 children). The authors support the higher probability value - which still over-identifies several children - by arguing that these children were also likely to have lower language abilities and may benefit from support.

Sachse et al. (2007) used critical values for language comprehension and production at 12 months on the ELFRA-1 in order to classify children at-risk for language delay as introduced above. Children were classified as at-risk if they scored below the critical value in at least one category. At 24 months, the ELFRA-2 classified children as late-talkers if they produced less than 50 words in spontaneous speech. The classification agreement between the ELFRA-1 and the ELFRA-2 was described by a 2 x 2 contingency table indicating a sensitivity (i.e. the proportion of late talkers children correctly identified as at-risk) of 52% and a specificity (i.e. the proportion of typical children correctly identified as typical on the ELFRA-1) of 65%. Furthermore, the RIOC-Index (‘Relative Improvement over chance‘) was 23% (RIOC < 34% = insufficient predictive power). According to the RIOC-index, prediction increased with the change of criterion at 12 months (below critical value for comprehension: RIOC = 39%; below critical value in comprehension and production: RIOC = 53%). However, the sensitivity was lower for both (i.e. 43% and 38%) but the specificity increased. Again, this means that the predictive validity of the ELFRA-1 is insufficient. It should be noted that predictive validity improved when taking into account male gender and low maternal education using group mean difference tests, but the authors advised to be cautious with the results due to small sample sizes. However, if the authors had decided to use a discriminant analysis, it would have been possible to examine how well these factors help to classify participants into groups also in regards to their sample size.

Westerlund et al. (2006) used receiver operating characteristic (ROC) curves. ROC curves were used to plot all possible cut-off points for the predictor variables of the Swedish CDI (SCS18) at 18 months (predictor variables: word production, comprehension, gesture, combination of the three variables) in order to decide which sensitivity and specificity values were the best when compared to the formalized language observation of receptive and expressive language (LO-3) by a child health care nurse for severe delay at 3 years (Zweig & Campbell, 1993). Production at 18 months on the SCS18 was the best predictor on the ROC curve with the criterion for language delay of less than 8 spoken words. Logistic regression also confirmed that production was the only significant predictor for the outcome of the LO-3 (Cox & Snell R Square or Nagelkerke R Square were not reported to estimate the variation explained). The sensitivity was 50% and the specificity was 89.6% (positive predictive value = 89.6%, negative predictive value = 97.6%, likelihood ratio (CI 95%) = 4.8 (2.7 – 8.4)). Similar to the studies above, many children (50%) with a severe delay at 3 years could not yet be reliably detected at 18 months using the SCS-18.

Unhjem, Eklund and Nergård-Nilssen (2015) used logistic regressions to test if knowledge of risk status (familial risk for dyslexia vs controls) led to better predictions for language delay at 24 months if tested on the Norwegian CDI at 18 months (CDI:WG receptive and productive vocabulary). It was found that children from both groups were equally likely to have language delay at 24 months (< 20^{th} percentile for production on CDI:WS). For the at-risk group productive vocabulary at 18 months was a significant predictor and the model correctly identified 9 out of 11 late-talkers (sensitivity 81.8%), and 18 out of 21 who were not late-talkers (specificity 85.7%). There was a strong relationship between prediction and the two groups (Nagelkerke's R Square = .61; Odds ratio = .08). On the other hand, for the control group the model could not identify children with language delay at 24 months (i.e. CDI:WG comprehension and production at 18 months were not significant predictors). It should be noted that the sample size was quite small (at-risk N = 32, control N = 21).

The results of this study show that whilst late-talker status could not be predicted between 18 and 24 months for typical children, the early language was more stable for at-risk children which allowed a significant prediction of language status at 24 months.

3.10. Summary and Conclusion

The predictive validity of CDI:WG questionnaires has been analysed for a range of languages, age groups, different subscales of the CDI and follow up instruments and also in terms of group-level or individual-level analyses. Looking at the group-level data, it becomes clear that the gesture, comprehension and production subscales have been widely used for predictive value research, whereas the Phrases Understood scale has been mostly disregarded.

3.10.1. Gestures

The literature reviewed above has shown that gestures were stable during infancy as the correlations were significant and moderate in strengths (Bavin et al., 2008; Fenson et al., 1994; Fenson et al., 2007). Gestures have also shown to be a precursor for future language ability. From 12 to 13 months, children learn to direct the attention of others through communicative gestures (Colonnesi, Stams, Koster, & Noom, 2010; Igualada et al., 2015). The first type of gesture used is usually the communicative pointing gesture which forms part of the group of deictic gestures including showing and giving. During the second year, children also start to use symbolic gestures (CDI categories: Actions with Objects, Pretending to be a Parent, Imitating other Adult Actions, Pretending Objects). The transition from functional (i.e. involves the pure function of objects, for example put cup on saucer) to symbolic play (i.e. the use of objects or actions to represent other objects or actions in play) reflects an increase in cognitive ability and attentional (including joint attention) skills and the variability in skills is also dependent on the quality of the child-caregiver relationship (Beeghley, 1998).The studies presented here mainly, but not exclusively used the communicative gesture subscale in order to predict future language comprehension and production. Communicative and symbolic gestures showed moderate relationships with comprehension during the early stages of communication (between 8 and 12 months). From 12 months, communicative gestures correlated weakly with later comprehension scores during toddlerhood and not at all during the preschool-period. Symbolic gestures also showed only weak correlations with comprehension during the toddler period.

More studies have so far looked at the relationship between early gestures and later language production. Other research has shown that communicative pointing preceded and even predicted future language ability (Colonnesi et al., 2010; Goldin-Meadow, 2007; Lueke, Ritterfeld, Grimminger, Rohlfing, & Liszkowski, 2015; Suttora & Salerni, 2012). Multiple studies have found that communicative gestures including pointing but also other gestures (e.g. conventional (i.e. gestures usually understood by one’s own culture, for example hand waving to say hello), representational gestures (i.e. gestures in place for a word, for example flap arms for bird) at around 14 months help to predict future language (Iverson & Goldin-Meadow, 2005; Rowe et al., 2008; Rowe et al., 2012).

The current literature confirms these findings as predominantly communicative gestures correlated moderately and predicted language production up to 24 months. However, the most long-term correlation during the infant-toddler period from 8 to 24 months showed only a weak association – possibly due to the neural reorganization which takes place around the child’s first birthday (McCall, Eichorn, Hogarty, Uzgiris, & Schaefer, 1977). Nevertheless, the strengths of association between gestures and production was somewhat stronger than with comprehension. Communicative gestures showed no significant relationships with language from 36 months of age. Yet, it was found that the effect of gestures at 15 months on risk for language impairment in the preschool period was mediated by production skills at 15 months (Hsu & Iyer, 2016). This shows that the impact of early gestures on later language does not stop after 24 months but after that age may be more important for less skilled language users, this was also supported by the study from Thal et al. (1997).

3.10.2. Phrases Understood

This subscale does not seem to carry the same importance for researchers as the vocabulary checklist, as it has been frequently omitted in shortened versions (CDI Short form) or alternative CDI adaptations (Oxford CDI, ELFRA-1). In this literature review only two studies incorporated this scale.

Phrases Understood correlated weakly to moderately with future language. Furthermore, language production could be predicted better than grammar or comprehension if first tested at 12 and 15 months and followed up at 2 and 4 years of age. Phrases Understood at 15 months together with child behavioural variables accounted for some significant variance in (expressive and receptive) language scores at 4 years. This shows that this category gives valuable information when predicting future language skills and thus should be incorporated when analysing the predictive validity of CDIs.

3.10.3. Comprehension

There was stability in language comprehension during the infant and toddler period. Generally, the correlations were moderate in strengths and early comprehension could explain some of the variance in later comprehension scores. Early vocabulary comprehension also contributed significantly to language comprehension in the preschool period (Fish & Pinkerman, 2003).

In early language development comprehension precedes and predicts production (Bleses, Vach et al., 2008). Around the child's first birthday a higher variability in scores is expected in vocabulary comprehension compared to production. Early comprehension may be a stronger predictor than early production for later language. Overall, there was a moderate relationship between vocabulary comprehension and later production.

3.10.4. Production

Early production ability was much better than comprehension in explaining some of the variance in future production skills or overall language ability. Furthermore, early production helped to predict later risk for language impairment (Hsu & Iyer, 2016). The relationships between early and later production were generally moderate to strong. Thus, the results of this review show that production was most stable out of the categories examined.

Only a few studies investigated the long-term association between early production and later comprehension. The results were not very consistent across studies as these varied between no significance, or small to moderate relationships. There were also some mixed results in terms of early production as a predictor for later comprehension. Therefore, it was suggested that other preverbal skills such as gestures, cognitive abilities (see Rose et al., 2009 for information processing), but also other external and internal factors, for example early maternal responsiveness and child intentional communication (Paavola et al., 2005), may be better predictors for comprehension during the infant and toddler phase.

Overall, follow-up at 4 years or older usually showed no or only small significant correlations between early word production and later language (i.e. word production, grammar, comprehension and total language scores) except for Hamadani et al. (2010). However, several studies found that early communication (comprehension, production, gestures) made significant contributions to language and literacy outcomes during the preschool and primary school phase (Duff et al., 2015; Fish & Pinkerman, 2003; Hsu & Iyer, 2016).

3.10.5. Summary: Group-level analyses

The studies demonstrated that a set of different variables is better at explaining the variance in future language ability than using single predictors. Several studies have used CDI categories (comprehension, production, gestures or Phrases Understood) separately with other variables, for example age, SES, maternal sensitivity, other child variables etc. (e.g., Fenson et al., 1994; Fish & Pinkerman, 2003; Laakso et al., 1999) which together helped to explain up to 51% of variance in the outcome variable with predictor variables assessed several months before the outcome variable (P. Lyytinen et al., 1999).

Creating latent variables/composite scores/several language variables used in regression analyses created from different CDI scores (gestures, Phrases Understood, comprehension, production) can help to create a better picture for prediction (e.g., Duff, Reen et al., 2015; Rose et al., 2009; Thal et al., 1997). For example, the Computerized Comprehension Task (CCT), an in-person assessment, was better at predicting language use when the CDI and CCT measures were directly compared for their predictive value (Friend et al., 2012). However, only the CDI's language comprehension scale was taken into account for this study. It is possible that composite scores (e.g. comprehension AND production) would have yielded a higher predictive value as found in studies which looked at such paired factors (e.g., Berglund & Eriksson, 2000; Corkum & Dunham, 1996).

Interestingly, Thal et al. (1997) looked at a similar variable (percentage of all words understood that were also produced) and found that this variable and vocabulary comprehension differently affected the future status of late and early talkers. Furthermore, the researchers investigated different ability levels on the CDI. However, they did not create categories depending on low comprehension and low gesture scores at 13 months in their study using the US CDI norm sample, but instead predicted only from precocious talkers at that age. Due to floor effects for production scores at this age, a late-talker category was not created. As there is more variability in comprehension and gesture scores at 13 months, it would have been interesting to find out if those low scores could significantly distinguish between late, average and early talkers at 20 months using the big norm sample (N = 217).

In contrast, another study by Thal et al. (2013) investigated the trajectory of three groups, one typical and two delayed groups: late-talkers who were only delayed in terms of production and late talkers who were also delayed in terms of comprehension. Late talkers with comprehension delay had the slowest development from 10 months to 5 years when compared to late talkers with no comprehension deficit and typically developing children. However, the two delayed groups caught up with the typical children on most language and vocabulary tests at 4 to 5 years of age. At 5 years, there were also no differences between the three groups in terms of cognitive (K-ABC) as well as social, self-help, gross or fine motor outcomes (Minnesota CDI). However, the delayed children showed significant delays on some measures of phonological working memory. This is important as deficits in phonological processing are a recognized marker for developmental dyslexia (e.g., de Carvalho, Kida, Capellini, & de Avila, 2014). There is the notion of illusory recovery which denotes that even though a high percentage of late talkers is known to catch up and score within the normal range on language tests during school age, they are still at elevated risk for literacy problems (Scarborough & Dobrich, 1990); however, this point of view is not shared by all researchers (e.g., Dale et al., 2014). Furthermore, CDI communication (comprehension, production and gestures) tested at 16 months as well as early growth of vocabulary comprehension from 10 to 16 months were important predictors for language status (production and/or grammar delay versus no delay at 28 months).

An interesting finding was that associations between early language, particularly comprehension (below 15 months) and later production were a lot weaker for the control children (Unhjem et al., 2015) compared to other studies which did not exclude those children at-risk for dyslexia (e.g., P. Lyytinen et al., 1999). The study by Lyytinen et al. (1999) analysed at-risk children (50% of the sample) together with data of typical children. Their multiple regression analysis predicting vocabulary production at 24 months from variables assessed at 18 months (i.e. play, language, gender and maternal education) was able to explain the highest variance (51%) of all studies compared here, except for Fish and Pinkerman (2003) who assessed some child behaviour predictor variables and the outcome variable at the same time. Lyytinen et al. (1999) found that only verbal comprehension (RDLS) and vocabulary production significantly contributed to the overall variance. Thus, it can be suggested that the high number of at-risk children may have elevated the prediction which would have potentially shown less continuity if the sample had been more representative of the population.

Future research should assess if there is more stability in early language for specific subgroups as suggested by some studies, for example for children at-risk of dyslexia (Unhjem et al., 2015), impoverished children from poor areas (Hamadani et al., 2010), or children with lower maternal education (Reese & Read, 2000). This means the stability of language production but also early comprehension as an indicator for later language should be examined for these subgroups. Possibly, the link for early vocabulary (comprehension and/or production) during infancy and later production may be even less strong as suggested by the studies here if risk factors/subgroups were removed.

The overall results of the group-level analyses suggest that generally the associations were stronger if the same linguistic domain (e.g. communicative gestures or comprehension or production) was used at follow up testing. In line with these findings, early production scores showed somewhat stronger relationships with future vocabulary production compared to grammar scores (Feldman et al., 2000; Kreisman, 2012; P. Lyytinen et al., 1996). Furthermore, stronger associations were found if the same instrument (P. Lyytinen et al., 1996) or at least category (i.e. vocabulary, see Duff, Reen et al., 2015) was used at follow-up. The associations were usually stronger if the interval between the tests was shorter. In many cases, the strengths of the relationship was stronger if age at time 1 was later (i.e. towards the end of the CDI:WG age span).

Most studies reviewed here looked at group-level analyses, therefore conclusions drawn here maybe more reliable than for individual-level analyses as their use was very limited.

3.10.6. Summary: Individual-level analyses

For the individual-level analyses, classification accuracy values were calculated in order to show how well the CDI could be used to correctly identify who would go on to develop a language problem.

However, there is a problem with the comparability of the results of the studies, as researchers reported different types of classification accuracy values which cannot always be converted to the same values as information is missing. Specifically, in some studies only the overall classification accuracy was reported without giving information about specificity and sensitivity values (Fish & Pinkerman, 2003; Thal et al., 1997). To compare results across studies, Dollaghan (2013) suggested to report confidence intervals (CI) to estimate if the accuracy value obtained is close to the true accuracy value as well as likelihood ratios (LR). These use the likelihood of a positive or negative screen, expressing the odds that a child would be correctly identified using a given cut-off criterion of a screening test. However, no CIs were reported and LRs have only been reported by one study (Westerlund et al., 2006). Likelihood ratios can be calculated as long as the specificity and sensitivity values or raw scores are known (Sachse, Saracino et al., 2007; Unhjem et al., 2015).

Specificity was generally higher than sensitivity, this was also found in a systematic review of speech and language measures for older children (Law, Boyle, Harris, Harkness, & Nye, 2000a). However, for clinical purposes the aim is to reliably detect those children who will go on to develop language problems (i.e. high sensitivity) even at the cost of potentially over-identifying some children without later language difficulties (i.e. reduced specificity) as suggested by Westerlund (2006).

The optimal accuracy values should be at or higher than 80% for sensitivity and specificity (Westerlund et al., 2006), but there are no exact guidelines to follow for speech and language measures (Law, Boyle, Harris, Harkness, & Nye, 2000a). For the studies reviewed here, the sensitivity levels were around 50% for typical children (Sachse, Saracino et al., 2007; Westerlund et al., 2006) and acceptable for the at-risk group for dyslexia (Unhjem et al., 2015). Furthermore, the overall correct classification showed that between 1/3 and 1/4 of children could not be correctly identified (Fish & Pinkerman, 2003; Thal et al., 1997).

Apart from assessing the predictive validity of representative populations, future research should also examine specific subgroups more (as mentioned above). As it is common practice to create norm scores separately for boys and girls, it may be suitable to create language growth curves separately for different subgroups, for example as a function of birth weight (Stolt et al., 2009), SES (Reese & Read, 2000) or poor areas (Hamadani et al., 2010). This may be especially useful for clinical purposes, as individual children with certain characteristics are more at risk for persistent language difficulties and the study by Unhjem (2015) has already found acceptable classification values for the at-risk group. Furthermore, the incorporation of additional risk or protective factors may also be desirable in order to increase the predictive power.

3.10.7. Conclusion

On the group-level, most studies followed children up to the toddler or preschool period and found that early communicative skills were associated with and made significant contributions to later language. The extent of stability depended on the category, production showed the strongest stability over time.

On the individual-level, the CDI:WG did not show satisfying predictions for individuals if followed up between 7 and 49 months later. However, there is some indication that classification accuracy may be improved if established for separate subgroups. Ideally, future research should investigate the language trajectories and predictive values separately for different at-risk populations.

4. The current longitudinal study

This PhD project assesses the predictive validity of the newly adapted UK-CDI questionnaire (see Chapter 2.3., Alcock et al., in prep) between 12 and 36 months. This PhD is a longitudinal project which includes typically developing children from a range of socio-economic backgrounds from the East Midlands (UK) with no known clinical or developmental disorders. It is essential to have appropriate assessment methods with long-term reliability to identify children at risk of language delay early in life. However, the most commonly employed UK language assessments for young children have so far not yet been assessed for their predictive validity. For example, the Ages and Stages Questionnaire (ASQ, see Squires, Twombley, & Bricker, 2009) is currently in use as a population-based screening tool for children between 24 and 30 months to detect developmental delay. This parent-report questionnaire examines all areas of development but includes very few questions about communication and language. As yet, the ASQ does not provide UK norms and no predictive validity data for British children (Bedford, Walton, & Ahn, 2013; Velikonja et al., 2016).

Standardised tests are a different tool available for UK infants and commonly used by Speech and Language Therapists if a child has been referred to them due to slow or atypical language development. The gold standard for British English assessments are the UK Preschool Language Scales (Zimmerman et al., 2014) for language and the Bayley Scales of Infant and Toddler Development (Bayley, 2010) for cognitive, motor and language skills. While these tests have been standardised, they afford lengthy and time-intensive in-person testing with the child and, again, have not been assessed for their predictive validity by the test developers.

While we cannot assume good predictive validity for the UK-CDI from the findings of other CDIs, other research methods have also found good predictability during early communication development. For example, observational research during parent and child interactions (e.g., Iverson & Goldin-Meadow, 2005; Rowe et al., 2008) ascertained that gestures were associated with later vocabulary comprehension and production; a pattern which was also found in the above mentioned CDI studies (Bavin et al., 2008; Feldman et al., 2000; Kreisman, 2012; P. Lyytinen et al., 1996; Sachse et al., 2007) with few exceptions (Fish & Pinkerman, 2003; Rose et al., 2009). This supports the view that some prediction is possible but we still require an instrument which can be used for this purpose.

The UK-CDI was the tool of choice for this project as it has been standardised and it is a normed parent-report language and communication questionnaire (Alcock et al., 2017). It can be especially valuable for researchers and clinicians alike due to its cost-effectiveness and ease of administration and measurements. It is also the first standardised CDI for British-English speaking children and it is based on a representative sample of children across the UK's regions and population, including data representative for gender and ethnic background. This newly developed parent-report instrument is for children between 8-18 months and focuses on communication, in particular on receptive and expressive vocabulary and gestures. While the psychometric properties of the newly standardised UK-CDI have been examined for internal consistency, reliability and validity as described above (see Chapter 2.3.), there is no data about how well we can predict children's later language outcomes from early UK-CDI scores. Such data would be important and useful for early years practitioners and clinicians for use in daily practice, and for researchers to use in empirical studies on children's language acquisition and language delay. Whilst the predictive validity of other versions of CDIs:WG have been assessed (see previous chapter) and good predictability could be generally shown during the early years, we cannot automatically assume that the UK-CDI yields the same results. This is because direct comparison between cultures may be problematic, for example, as parental interaction styles across cultures may differ (e.g., Greenfield, Keller, Fuligni, & Maynard, 2003; Weber, Fernald, & Diop, 2017) or language-specific differences may lead to different rates of language development (Bleses, Vach et al., 2008; Bleses, Basboll, Lum, & Vach, 2011). For example, Hamilton et al. (2000) found that differences exist between US American children's language learning and the Oxford-based samples of British children with American children showing higher CDI scores.

For British English, attempts to investigate predictability from non-normed data sets using the Oxford CDI (Hamilton et al., 2000) have been made, mostly to study late talkers and their later language abilities, but also to predict children's school-age language and literacy from infant vocabulary (Duff, Nation, Plunkett, & Bishop, 2015; Duff, Reen, Plunkett, & Nation, 2015; Cochet & Byrne, 2016). Results suggest that there is a strong relationship between early vocabulary comprehension and later vocabulary comprehension and production as well as between early production and later production between 1 and 3 years (Cochet & Byrne, 2016). However, children classified as late talkers in terms of expressive vocabulary at 18 months fell within the normal range and did not differ any more from typical children at 7 years in terms of language and literacy skills, although, classification (SLI versus typical children) at 4 years showed good stability when tested again at 7 years (Duff et al., 2015). Whilst it was not possible to predict language outcomes on an individual level, group level analysis showed that infant vocabulary (16-24 months) significantly predicted later vocabulary for children aged between 4 and 9 years. When taking family risk for language and literacy difficulties into account, the prediction was even better particularly for reading outcomes (i.e. infant vocabulary explained 30% of variance in later reading comprehension) as children with a family risk and small vocabularies were more likely to develop reading problems (Duff, Reen et al., 2015). However, the data was gathered using the Oxford CDI which is described as a translation of the MB-CDI. It leaves out the gesture scale and in one instance has been used for a different age range than advised. Furthermore, it has not been standardised and it has no valid UK-wide norms.

The newly standardised UK-CDI offers the chance to overcome these problems. The UK-CDI in combination with other tools enables us to investigate the predictability of gesture, comprehension and production in typically-developing children and compare children's results to valid population norms.

If the UK-CDI has good predictive validity up to 2 or 3 years, children with language delay could be identified earlier. It is possible that the standardised UK-CDI could be used as a population-based tool to determine language status, similar to the German FRAKIS (Szagun, Stumper, & Schramm, 2009). Currently, no other detailed and at the same time cost-effective and easy to administer tool for determining language status exists. So far, the UK-CDIs predictive validity has not yet been assessed. Hence, this PhD project plans to close this gap in the literature.

The aims of the current study are as follows. The UK-CDI's short-term (12 - 18 months; 18 - 24 months) and long-term (12 - 24 months; 12 - 36 months; 18 - 36 months) predictive validity will be examined for the first time. This PhD research reports the predictive validity of the UK-CDI using a UK sample of typically developing children. It was beyond the scope of this longitudinal research to follow the trajectories for children at-risk for language problems, hence we limited ourselves to a neurotypical sample.

Predictive validity will be assessed in terms of language level. It will be examined if four groups of children (i.e. low (1-25th percentile), low-medium (26th-49th), medium-high (50th-74th) and high level (75th-99th)) remain stable over time or whether predictions are better for some groups than for others. This corresponds with the study by Duff et al. (2015) who also used the 25th percentile as the cut-off for the late-talking group at 18 months.

Research has shown that language scores of early talkers between 13 and 20 months as well as early and late talkers between 20 and 26 months were stable; however, the stability of language was somewhat stronger for early talkers (Thal et al., 1997). The current study will extend the study by Thal et al. (1997) by assessing different age groups (12, 18, 24 and 36 months).

There are several CDI: Words and Gestures studies which predicted language up to 36 months or even later (see previous chapter). At this stage most children are confident users of a large vocabulary and utter sentences of increasing grammatical complexity. Therefore, many studies use standardised language measures at the follow-up stage. These usually measure receptive and expressive language but do not give an indication of vocabulary ability alone. The current study used CDI-type questionnaires at all age points (12, 18, 24 and 36 months) as a measure of vocabulary (amongst other categories) which allows the direct comparison of vocabulary over time.

In addition, CDIs measure additional age-appropriate skills (e.g. gestures during infancy, grammar during toddlerhood). Therefore, it will also be possible to establish if later acquired skills using further assessment tools (i.e. CDIs, PLS-5 UK and ASQ-3) can be predicted from early communicative skills recorded on the UK-CDI.

Apart from language skills, the current study also measures the effects of other factors. In contrast to most other CDI predictive validity studies, this study includes background information about the child and family (e.g. socio-economic data, child's health status). The current study also includes other areas of development and more detailed child and family characteristics using the UK-CDI Family Questionnaire (e.g. prematurity, gender, SES, sibling status, ear infections, sleep, family risk of dyslexia or speech or language problems) to aim for a more holistic representation of child development over time and factors influencing children's language outcomes.

As cognitive processing (e.g., Rose et al., 2009) and motor development (e.g., Iverson, 2010; Leonard & Hill, 2014) are associated with language development, these skills were measured as well to assess how early language and other factors influencing language development help to predict outcomes at 2 and 3 years. Taking into account these different factors, this will show if language development differs depending on subgroups (Reese & Read, 2000).

As the UK-CDI is a new instrument which may have effective practical use, it is important to evaluate its classification accuracy. It will be examined which children with low UK-CDI scores at 12 or 18 months will go on to have delayed language at 24 or 36 months.

Based on the review of the previous studies using CDIs surveyed in Section 3, the predictive outcomes of the UK-CDI are likely to be the following:

1) We expect stronger associations if the same linguistic domain is tested at follow-up (e.g. production scores to correlate with production scores at the next stage).
2) We also assume stronger associations with another CDI measure at follow-up compared to a different instrument used at the same age (for example CDI:WS versus PLS-5 UK)

3) We expect stronger associations between UK-CDI:WG and later scores if the time interval is shorter (for example: stronger correlations between 12 and 18 months compared to 12 and 36 months)
4) Furthermore, as language is more established at 18 months compared to 12 months, we assume stronger correlations of UK-CDI:WG scores at 18 months with later language skills.
5) We expect that the language of high ability children as classified on the UK-CDI:WG to remain more stable over time compared to those children with low language ability
6) The individual prediction for later language status is expected to be difficult but may be improved with modified criteria. In this study, it will be assessed if predictions for later language delay (at 24 or 36 months) can be improved by using higher cut-off scores of early delay (at 12 or 18 months) than commonly employed by researchers.

5. ANALYSIS 1: RELATIONSHIPS BETWEEN UK-CDI SCORES AND LATER LANGUAGE SCORES

5.1. METHOD

5.1.1. Participants

At the start of the study, 125 parents with their children were recruited and took part in Test Time 1 (at around 12 months). The data from 39 children were not included in this analysis as parents only completed the CDI questionnaire either at Time 1 (N=13), at Time 1 and Time 2 (N=13), at Time 1, Time 2 and Time 3 (N=8) or at Time 1 and Time 3 (N=5). The reason for sporadic or discontinued participation was moving to a new house, parents did not respond to participation requests or they were too busy. In addition, some children were excluded from the study due to hearing another language more than 10 hours per week (N=3) or being deaf in one ear (N=1).

The final data set in this longitudinal cohort included the same 82 children at four timepoints (see Table 15 for detailed information). The children who were included in the research (N = 82) did not differ significantly from the children who were excluded (N = 43) in terms of sex ($\chi^2(1, n = 125) = 2.17, p = .14$, phi = .15), maternal age ($\chi^2(4, n = 124) = 1.25, p = .87$, phi = .10), maternal education ($\chi^2(4, n = 123) = 5.80, p = .21$, phi = .22) or family household income ($\chi^2(3, n = 122) = 1.16, p = .76$, phi = .10).

The age spread of this remaining cohort was widest for Time 1 (10 - 16 months) which was due to the age of when the children were recruited. The mean age was 11.71 months (*SD* 1.19) at Time 1, 17.83 months (*SD* .64) at Time 2, 24.46 months (*SD* .44) at Time 3 and 36.74 months (*SD* .60) at Time 4, see Table 15 for information about gender. The children were all from the East Midlands, UK. Over half the children were recruited in-person (e.g. baby groups at Children's Centres, libraries, community centres). Other participants were recruited via referral from a friend, University's internal communications, the Lincoln Babylab database and recruitment via social media.

Table 15. *All children included in longitudinal cohort separated by gender and age*

Time*	0;10	0;11	1;0	1;1	1;2	1;3	1;4	1;5	1;6	1;7	1;11	2;0	2;1	3;0	3;1	3;2
Time 1																
girls	2	14	12	3	1	1										
boys	3	27	9	4	4	1	1									
total	5	41	21	7	5	2	1									
Time 2																
girls							1	8	21	3						
boys								14	28	7						
total							1	22	49	10						
Time 3																
girls											2	25	6			
boys											3	41	5			
total											5	66	11			
Time 4																
girls														27	5	1
boys														32	14	3
total														59	19	4

*Time = time point at completion of CDI

All participants were singletons and of white British ethnicity. None of the children were exposed to a second language. The primary carer(s) was the mother or mother and father except for one child who lived with father full-time, see more details for marital status in Table 16. All participants lived more than half the year with the parent who completed the questionnaires of this study. None of the children were diagnosed with a developmental disability or visual/hearing impairments. One child had hypertonia hypermobility. Three children had two or more ear infections up to 18 months of age. Further participant characteristics are given in Table 16 below.

Table 16. *Participant characteristics*

	Boys (n=49)		Girls (n=33)		Total (n=82)	
Characteristic	n	%	n	%	n	%
Maternal characteristics						
Maternal age (at age 12m)						
Up to 20 years old	1	2.0	0	0.0	1	1.2
21 to 25 years old	3	6.1	3	9.1	6	7.3
26 to 30 years old	18	36.7	11	33.3	29	35.4
31 to 35 years old	18	36.7	13	39.4	31	37.8
36+ years old	9	18.4	6	18.2	15	18.3
Maternal education						
No formal education	1	2.0	0	0.0	1	1.2

GCSE/O Level/NVQ Level 1 or 2/similar	3	6.1	3	9.1	6	7.3
A Level/NVQ Level 3/similar	9	18.4	12	36.4	21	25.6
University degree/HND/HNC/ NVQ Level 4 or 5/similar	21	42.9	13	39.4	34	41.5
Postgraduate/similar (PGCE, PhD, MA etc.)	15	30.6	5	15.2	20	24.4
Paternal characteristics						
Paternal age (at age 12m)						
Up to 20 years old	0	0.0	0	0.0	0	0.0
21 to 25 years old	0	0.0	4	12.1	4	4.9
26 to 30 years old	15	30.6	8	24.2	23	28.0
31 to 35 years old	13	26.5	10	30.3	23	28.0
36+ years old	21	42.9	11	33.3	32	39.0
Paternal education						
No formal education	0	0.0	0	0.0	0	0.0
GCSE/O Level/NVQ Level 1 or 2/similar	11	22.4	4	12.1	15	18.3
A Level/NVQ Level 3/similar	15	30.6	16	48.5	31	37.8
University degree/HND/HNC/ NVQ Level 4 or 5/similar	13	26.5	10	30.3	23	28.0
Postgraduate/similar (PGCE, PhD, MA etc.)	10	20.4	3	9.1	13	15.9
Parental characteristics						
household income						
0 to 14000*	4	8.2	0	0.0	4	4.9
14001 to 24000	5	10.2	4	12.1	9	11.0
24001 to 42000	22	44.9	12	36.4	34	41.5
42001 or more	18	36.7	16	48.5	34	41.5
Missing	0	0.0	1	3.0	1	1.2
Bedrooms						
2 bedrooms	9	18.4	3	9.1	12	14.6
3 bedrooms	27	55.1	19	57.6	46	56.1
4 bedrooms	10	20.4	10	30.3	20	24.4
5 + bedrooms	2	4.1	1	3.0	3	3.7
Missing	1	2.0	0	0.0	1	1.2
Marital Status (at 12 months)						
Married/Civil Partnered	28	57.1	26	78.8	54	65.9
Living with partner	18	36.7	7	21.2	25	30.5
Single	3	6.1	0	0.0	3	3.7
Disorders in first degree relatives						
Autism	0	0.0	0	0.0	0	0.0
Speech and Language Difficulty	12	24.5	7	21.2	19	23.2
Dyslexia	7	14.3	5	15.2	12	14.6
Child characteristics						
Gestational age at birth (weeks)						
Week 34 - 36	6	12.2	2	6.1	8	9.8
Week 37 or later	43	87.8	31	93.9	74	90.2
Birth weight						
Up to 5lb 8oz	2	4.1	2	6.1	4	4.9
5lb 9oz to 9lb 14oz	45	91.8	30	90.9	75	91.5
9lb 15oz or over	2	4.1	1	3.0	3	3.7

Siblings						
No older siblings	34	69.4	16	48.5	50	61.0
1 older sibling	10	20.4	14	42.4	24	29.3
2 older siblings	4	8.2	1	3.0	5	6.1
3 or more	1	2.0	2	6.1	3	3.7
Sleep at 12 months						
9 hours	2	4.1	1	3.0	3	3.7
10 hours	0	0.0	0	0.0	0	0.0
11 hours	4	8.2	1	3.0	5	6.1
12 hours	8	16.3	7	21.2	15	18.3
13 hours	14	28.6	6	18.2	20	24.4
14 hours	19	38.8	12	36.4	31	37.8
15 hours	2	4.1	1	3.0	3	3.7
Missing	0	0.0	5	15.2	5	6.1
Sleep at 18 months						
9 hours	0	0.0	1	3.0	1	1.2
10 hours	0	0.0	0	0.0	0	0.0
11 hours	7	14.3	2	6.1	9	11.0
12 hours	14	28.6	5	15.2	19	23.2
13 hours	16	32.7	14	42.4	30	36.6
14 hours	8	16.3	8	24.2	16	19.5
15 hours	2	4.1	1	3.0	3	3.7
Missing	2	4.1	2	6.1	4	4.9
Childcare (Childminder/Nursery) at 12 months/ Hours						
No	28	57.1	16	48.5	44	53.7
Yes	21	42.9	17	51.5	38	46.3
1 - 20 hours	9	18.4	8	24.2	17	20.7
21 - 35 hours	7	14.3	7	21.2	14	17.1
36 + hours	4	8.2	2	6.1	6	7.3
Missing	1	2.0	0	0.0	1	1.2
Childcare (Childminder/Nursery) at 18 months/ Hours						
No	24	49.0	14	42.4	38	46.3
Yes	25	51.0	19	57.6	44	53.7
1 - 20 hours	10	20.4	8	24.2	18	22.0
21 - 35 hours	11	22.4	8	24.2	19	23.2
36 + hours	4	8.2	3	9.1	7	8.5

Note here: * two out of four from the lowest income bracket came from single households

At 18 months, two parents reported concerns about the child's communication. At 36 months, 10 parents (of which 8 children were boys) reported a concern about their child's speech and language (concerns: slow language development = 6; hard for other people to understand child = 9; poor pronunciation = 7; stutters = 1). Six out of these 10 parents already sought professional advice or treatment from a specialist. Furthermore, 19 parents (23%) reported that a first degree relative had a speech and/or language difficulty or dyslexia of which seven parents also had a concern about their child's current speech and language. See Table 17 below for an overview of family and child medical or health problems and concerns about child's development.

Table 17. *ASQ questions about family and child medical or health problems and concerns about child's development*

Questions about child's development (ASQ)	Number of children (N = 82)
Concerns about not hearing well	*4*
Family deafness	*11*
Concerns about vision	*3*
*Medical or health-related problems in last few months**	*9*
*Concerns about behaviour***	*4*
*Any worries about child****	*8*

**Asthma (N = 2), ear infections (N = 2), small surgery (N = 1), squint (N=1), tonsillitis (N=1), chest infections and sleep apnoea (N = 1), food allergies (N = 1)*
***Not potty trained (N = 1), physical with other children (N = 1), tantrums and lack of communication (N = 1), potential Aspergers diagnosis (N = 1)*
****Speech (N = 2), stummer and stutter (N = 1), stubbornness (N = 1), food allergies (N = 1), lack of eye contact (N = 1), coping with social situations (potential Aspergers diagnosis) (N = 1), feet turn in (N = 1)*

5.1.1.1 Premature children

The percentage of premature births was 8.5% (N = 7; five boys) and all of these children were born between 34 and 36 weeks. As it is common practice, prematurity was adjusted by term-correcting the age up to 24 months old in the current sample (e.g., Bayley, 2006a; Zimmerman et al., 2014). In the following, it was checked if preterm children performed below average in terms of their cognitive and motor scores on the Bayley-III and the auditory and expressive communication scores on the PLS-5UK. It was found that one child's score fell 1 SD below the mean on the motor scale at 18 and at 24 months due to hypertonia hypermobility. This child also had below average language (below 1 SD) in terms of auditory communication measured by the PLS-5 UK. The other children fell within or above the average range (i.e. four children scored above the mean in at least one scale at either age) at 18 and 24 months.

Furthermore, it was investigated if preterm children significantly differed from term children in terms of their vocabulary measured by the CDIs (i.e. productive and expressive) at 12, 18, 24 and 36 months of age. The analysis was run three times with different control children matched for age as closely as possible and gender. The vocabulary scores did not differ between preterm and term children at any time, so the data of the preterm children was retained in the data set, see Appendix 3 for results.

5.1.2. Materials

The following assessment tools were used to assess language (UK-CDI:WG and Family Questionnaire (Alcock et al., 2017), Lincoln CDI:WS (Meints & Fletcher, 2011), PLS-5 UK (Zimmerman et al., 2014), 3-year parent report language measure (Dionne et al., 2003) and ASQ-3 British version (Squires et al., 2009). Motor and cognitive skills were assessed using the Bayley Scales of Infant Development, Third edition (Bayley, 2006b). The detailed description of the materials will be reported where appropriate in the text below. See Table 18 for an overview of measures used at different ages. Some questionnaires used here have not been standardised for British children which are the Lincoln TCDI:WS (Meints & Fletcher, 2011), the 3-year parent report language measure (Dionne et al., 2003) and the British version of the ASQ-3 (Squires et al., 2009).

Table 18. *Measures used at different ages separated by assessment type*

Age group	language	motor	cognitive
12 months	**UK-CDI:WG** (Phrases Understood, Comprehension, Production, Gestures)		
18 months	**UK-CDI:WG** (Phrases Understood, Comprehension, Production, Gestures) **PLS-5 UK** (Auditory Comprehension, Expressive Communication, Total Language)	**Bayley-III** (fine, gross motor, total motor score)	**Bayley-III**
24 months	**Lincoln TCDI:WS** (Comprehension, Production, Sentence Complexity) **PLS-5 UK** (Auditory Comprehension, Expressive Communication, Total Language)	**Bayley-III** (fine, gross motor, total motor score)	**Bayley-III**
36 months	**3-year parent report language measure** (Production, Sentence Complexity, Advanced Grammar) **ASQ-3 British version** (Communication)		

5.1.2.1. UK Communicative Development Inventory: Words and Gestures (UK-CDI)

The standardised and normed UK-CDI: Words and Gestures (Alcock et al., in prep) for children between 8 and 18 months was used to measure how many words and phrases children understand and produce and how many gestures they use.

Psychometric properties were assessed by the UK-CDI research team and summarized in the UK-CDI manual (Alcock et al., 2017). The UK-CDI has excellent internal reliability with Cronbach's alpha for word comprehension and production (both at .99) and for gestures (at .98). The correlations between age and the different subscales were significant and moderate (age and production, $r = .48$, $p < .001$) to strong (age and comprehension, $r = .67$, $p < .001$; age and gestures, $r = .79$, $p < .001$). The association between the subscales were also significant and moderate (production and gestures, $r = .56$, $p < .001$) to strong (production and comprehension, $r = .67$, $p < .001$; comprehension and gestures, $r = .78$, $p < .001$).

The inter-form reliability between online versus paper versions was assessed and the Pearson's correlations were .83 ($p < .001$) for comprehension, .59 ($p = .001$) for word production and .69 ($p < .001$) for gestures. Participants (Total N= 31) completed both forms with a gap of between 3 and 15 days. Half of the participants completed the online questionnaire first. Scores at time 1 and time 2 did not differ significantly from each other. The relatively low correlation for word production could be due to possible floor effects at this age (11-15 months).

Concurrent validity was also assessed using the Preschool Language Scale – 5th UK Edition (Zimmerman, Steiner, & Pond, 2014). UK-CDI comprehension scores correlated significantly with the PLS auditory comprehension score ($r = .413$, $N = 32$, $p = .003$) and the total score on the PLS ($r = .457$, $N = 32$, $p = .009$). UK-CDI production scores were also significantly associated with the PLS expressive language score ($r = .391$, $N = 32$, $p = .027$). Furthermore, an object selection task as employed by Fenson et al. (1994; 2007) was used to assess concurrent validity, in which children were presented with two objects and asked to pick the named object. UK-CDI comprehension correlated significantly with the object selection scores ($r = .413$, $N = 32$, $p = .019$) and UK-CDI production scores also correlated significantly with the object selection task ($r = .433$, $N = 32$, $p = .013$). A "gesture challenge" task similar to the one employed by Alcock et al. (2015) was also employed to assess concurrent validity. For this task, children were asked to act out certain gestures. UK-CDI Gesture scores correlated significantly with the total score on the "gesture challenge" task but only at the one-tailed level ($r = .344$, $N = 32$, $p = .027$). As the correlations with the gesture scale were not very strong a home validation was also conducted in which children's spontaneous gestures and gestures during the "gesture challenge" were videotaped. Here, stronger correlations were found between the UK-CDI total gesture score and the total number of different gestures produced during spontaneous play and the “gesture challenge” (Spearman rho = .48, N = 27, $p < .001$). In addition, communicative gestures produced during the home visit (both play and “gesture challenge” sessions) correlated significantly with the UK-CDI total communicative gesture score (Spearman rho = .65, N = 27, p < 001).

Scoring the data: Understanding of first words and phrases was divided into First signs of understanding and Phrases Understood. The first asked three Yes/No questions about the first words and phrases children may understand (scores for Yes = 1, No = 0). The latter asks questions (N = 28) about children’s ability to understand short phrases (scores for understands = 1, blank = 0). For the word list, each word’s score is either 1 or 0; a score of 1 denotes ‘understands’ or ‘understands + says’ and a score of 0 means a blank response which denotes does not yet understand and say. Production scores are the sum of all words ticked as ‘understands + says’ (Min = 0, Max = 395). Comprehension scores are the sum of all words ticked as ‘understands’ or ‘understands + says’ (Min = 0, Max = 395). Percentage scores are calculated from full scores (i.e. count of all UK-CDI words). The gesture score is the sum of scores (Min = 0, Max = 63) of all five gesture categories (i.e. first communicative gestures, games and routines, actions with objects, pretending to be a parent, imitating other adult actions).

5.1.2.2. Family Questionnaire

As part of the creation and standardisation of the UK-CDI: Words and Gestures (Alcock et al., in prep), the Family Questionnaire (FQ) was developed to give to participants in one parent pack with the UK-CDI: Words and Gestures (see Appendix 4). The current study follows this practice as the FQ asks about issues relevant to this study: the child’s health, sleep, family circumstances including SES information (e.g. maternal education, household income and number of bedrooms) as well as childcare information. Apart from few exceptions (e.g. question about hours of sleep), most questions have been frequently connected to language development and therefore are often included in CDI research (e.g., Fenson et al., 1994; Fenson et al.,2007). Some questions ask about sensitive information (e.g. household income), thus, participants were reminded at the beginning of the questionnaire that they could leave any questions blank which they did not want to answer.

5.1.2.3. Lincoln Toddler CDI (Lincoln TCDI)

The British Lincoln Toddler CDI (Meints & Fletcher, 2011) is a cultural and linguistic adaptation from the original US CDI: Words and Sentences (Fenson et al., 1994; Fenson et al., 2007). The Lincoln Toddler CDI has not been standardised and normed for the UK population. In addition, the authors have not yet published any information in terms of the psychometric properties of this instrument. The Lincoln TCDI questionnaire is for children between 16 and 30 months of age and includes questions about the child's vocabulary (receptive and expressive) and grammar (combining words, mean utterance length (in morphemes), use of words, use of suffixes, irregular endings, over-regularisations and Sentence Complexity). In contrast to the original US version, the Lincoln TCDI asks about the child's expressive and also receptive vocabulary in a checklist that includes 689 words. For details of the adaptation, see Appendix 5.

Scoring: The Toddler CDI scores production and comprehension in the same way as the UK-CDI (Min = 0, Max = 689). Percentage scores are calculated by dividing the sum of scores by 689 (i.e. count of all Lincoln Toddler CDI words). There are no scores for gestures as the Toddler CDI does not include this subscale.

5.1.2.4. 3-year parent report language measure

The 3-year parent report language measure (Dionne et al., 2003) was used in the TEDS study (Twins Early Development Study), was made for British English children at the age of three years and is a precursor of the American English CDI-III (Fenson et al., 2007). The questionnaire has not yet been standardised and normed for British children; however, a large dataset exists for twins from the TEDS study (Dionne et al., 2003). It comprises an expressive vocabulary checklist of 100 words. The concurrent validity has been assessed for this instrument with the McCarthy Scales of Children's Abilities Verbal Score (Oliver et al., 2002) as well as with the Preschool Language Scale -

3 (cited in Dale, Reznick, Thal, & Marchman, 2001) and moderate to strong correlations were reported (Dionne et al., 2003). No further psychometric properties have been reported by the test developers. Permission for questionnaire use at 36 months was granted by Philip Dale (University of New Mexico) via email on 6th December 2015.

Scoring: Parents are asked to tick words which they have heard their child say in spontaneous speech (production score: maximum = 100, minimum = 0). The next section is about grammar and the first question asks if the child combines words yet (answer types: often, sometimes, not yet). In addition, it contains 12 questions about frequency and grammatical complexity of word combinations. Parents are supposed to highlight which out of two utterances best reflects their child's current language use. Another 12 yes/no questions ask about general language use (e.g. understand and illustrate conceptual ideas and semantic abilities, talk about hypothetical events). In the following, we refer to the first grammar scale as Sentence Complexity scale and to the latter we refer to as Advanced Grammar. Scores of the Sentence Complexity scale are 1 for the more complex grammatical structure and 0 for the less complex sentence structure (maximum = 12, minimum = 0). For the Advanced Grammar scale, scores are 1 for yes and 0 for no (maximum = 12, minimum = 0). Furthermore, brief questions enquire about parental concerns about the child's speech and language and actions taken (if applicable). Six questions ask about the child's communicative behaviour to indicate potential communication abnormalities (e.g. showing no signs of joint attention skills, or a decline in language skills).

5.1.2.5. Ages and Stages Questionnaires (ASQ)

The Ages and Stages Questionnaires (ASQ-3, see Squires, Twombley, & Bricker, 2009) were used at 3 years as they aim to detect developmental delays or disorders. Permission for the online and paper use of the British ASQ (36 Month Questionnaire) was granted by Brookes Publishing Co. The questionnaires are available for children between 1 through to 66 months and have been translated and adapted from American English into British English in collaboration with the Department of

Health. The ASQ-3 is currently used as a population screening instrument for UK children between 24 and 30 months (i.e. integrated-2-year-review formerly known as Healthy Child Programme (HCP) 2-2.5 review), currently via a combination of health care and preschool routes.

The questions ask about communication (6 questions), gross motor (6), fine motor (6), problem solving (6) and personal-social development (6). Further, 10 yes/no questions (with space for explanations) enquire about the child's hearing, eye-sight, family history of deafness or hearing problems, health problems, talking, understanding and gross motor ability as well as parental concerns about behaviour or development.

There are currently no standardised UK norms available (Bedford, Walton, & Ahn, 2013), thus the British English version adopted the cut-off scores from the American English children. For this study the British ASQ 36 Month Questionnaire (34 months 16 days to 38 months 30 days) was used.

Whilst the psychometric properties of the British English version have not yet been assessed, the reliability (i.e. internal validity, test-retest reliability, inter-rater reliability) and diagnostic accuracy (sensitivity, specificity) of translated/adapted ASQ versions from American English into other languages (i.e. Spanish (Chilean), Portuguese (Brazilian and Portugal), Indian, Dutch, Korean, Turkish, Thai, Chinese) yielded mixed results (for a systematic review, see Velikonja et al., 2016). However, it should be cautioned that the comparison of psychometric properties using questionnaires in other languages may not be relevant, but it is the only available information we have at the moment.

Scoring: Possible answers are YES (score 10), SOMETIMES (score 5), NOT YET (score 0). Total scores (score range per subscale: 0 - 60) are created for each subscale which have age-appropriate cut-off scores (2 SD below the mean) and close-to-cut-off scores (between 1 and 2 SD below the mean) (Squires et al., 2009). This means children can be classified as delayed if scoring below 2 SD and those scoring between 1 -2 SD should be monitored according to the authors (Squires et al., 2009).

5.1.2.6. Preschool Language Scale 5-UK

The Preschool Language Scale 5-UK (PLS-5 UK) was used to establish the concurrent and long-term validity of the CDI instruments. This instrument provides assessment and norms for Auditory Comprehension and Expressive Communication on separate subscales. A total language score can be obtained by combining the standard scores of both subscales.

Language, even word to object associations, is difficult to test in young children, so the PLS-5UK makes use of behaviours which have shown to predict receptive and expressive future language (i.e. attention to sounds, objects and people; production of speech sounds; ability to communicate socially/interactively via gesture and play). As language develops children can be tested in terms of their word knowledge and grammar which is expected to increase in complexity over time (i.e. semantics: vocabulary/ connected speech, qualitative, quantitative, spatial and time/sequence concepts; language structure: morphology, syntax). The psychometric properties have been assessed by the authors (Zimmerman et al., 2014). The internal consistency was analysed for the UK version of the PLS-5 using split-half reliability. The Expressive Communication coefficients showed mostly good to excellent reliability ranging from .76 to .95, by age (.88 for overall normative sample) and similar for Auditory Communication ranging from .64 to .94, by age (.86 for overall normative sample). The total language composite showed an average reliability coefficient of .93 (ranging from .72 to .97 by age). The American PLS-5 (Zimmerman et al., 2011) showed excellent split-half reliability coefficients for language delayed children (.93 - .97) and children with a language disorder (.96 - .97). This demonstrated that the PLS-5 was not only a reliable measure for the general population but also for clinical groups. The manual also reported high concurrent validity results with the PLS-4 (Zimmerman, Steiner, & Pond, 2002) and moderate to high correlations with the CELF Preschool-2 (Wiig, Secord, & Semel, 2004) using the American PLS-5 but not the PLS-5 UK. Furthermore, the performance of children with receptive, expressive or mixed disorders differed significantly from matched peers on the PLS-5. Diagnostic Accuracy was established for the American PLS-5. For the Total Language score, a disorder was identified from 1

SD below the mean. It was found that the specificity (i.e. the probability that someone without the condition will test negative for it) was .80 and the sensitivity (i.e. the probability that someone with the condition will test positive for it) was .83. This meant that 83% of children who were previously diagnosed with a language disorder were also identified with a language disorder on the PLS-5 and that 80% of typically developing children were correctly classified as not having a language disorder on the PLS-5.

Scoring: Children with typical language should score between 1 SD below and 1 SD above the mean (standard mean score = 100).

5.1.3. Procedure

Parents who took part in this longitudinal study had previously expressed an interest for their child to take part in studies at the Lincoln Infant and Child Development Lab or they were purposefully recruited for this study by the experimenter at Children's Centres, Libraries etc. They were asked to complete and return the UK-CDI and Family Questionnaire when their child was aged around 12 and 18 months, the Lincoln Toddler CDI at 24 months and the 3-year parent report language measure and the ASQ at 36 months. At 18 and 24 months parents were also invited to take part in-person tests to assess children's language (PLS-5 UK, see Zimmerman et al., 2014) as well as motor and cognitive development (Bayley-III, see Bayley 2006b).

5.1.3.1. Procedure at 12 months

To start participation, parents of infants around 12 months were sent the UK-CDI and Family Questionnaire and asked to fill them in and return them to the Babylab. Parents could choose if they preferred paper or online completion and were sent the parent-report questionnaires according to their preference for the course of the longitudinal study. The aim was to recruit children around their first birthday (+/- 6 weeks). A few parents returned the questionnaires later than 6 weeks but were kept in the data set (see Table 15 above). The questionnaires were sent via an online link by email or via post including prepaid return envelopes, depending on parents' preference. If parents did not return the questionnaire within a couple of weeks, they were contacted again up to two times to remind them about the study. We stopped reminding parents once the child was too old (12 months + 6 weeks) even if they had received less than two reminders. To remind parents, we contacted them via telephone or via letter/ email depending on the contact information details given in the sign-up sheet. The reminders as well as the information letter in the parent pack were used to give more information about the longitudinal nature of this research which involved the future in-person assessments and questionnaires.

5.1.3.2. Procedure at 18 and 24 months

When their child was 18 and 24 months old, parents were contacted via email, letter or telephone two weeks before their child reached the appropriate age. If parents did not reply, they received a reminder via telephone call, email or letter with possible dates for the in-person meeting. If parents did not reply after the second request, it was assumed that they were not interested in taking part at that point. The sessions were arranged between the day of when the child turned the appropriate age until six weeks after that date. The questionnaire(s) was/were sent out one week before testing took place and had to be completed prior to taking part in the in-person assessment at the Babylab, early years' settings or at parents' homes depending on the parents' preference. These in-person sessions required the parent and child to be present and included language assessments using the PLS-5 UK as well as the use of the motor and cognitive subscales of the Bayley Scales of Infant and Toddler Development, Third Edition (Bayley-III) assessments.

At the beginning of the testing session, the caregiver and child were warmly greeted, and children were encouraged to play, and the researcher and parent had time to chat and get comfortable. Parents were asked if they were happy to be video recorded for the duration of the session. All parents agreed and signed a consent form before testing started. The Bayley Scales were administered in the same order (first cognitive, then motor scales) and before the PLS (see procedure of Bayley Scales in Analysis 3 below). This was decided in order to use the testing time efficiently as language and play behaviour could be observed during the first hour of using the Bayley which made the PLS scoring process faster. The PLS-5 UK was administered according to the manual. Total testing time was around two hours, but sessions were usually split into two within a few days apart.

5.1.3.3. Procedure at 36 months

Testing at 36 months involved completing only parent report questionnaires, and no in-person assessments were required. Parents received the parent pack (including the ASQ-3 British version and the 3-year parent report language measure) a few days before the child's 3rd birthday via email or post and were asked to complete and return the questionnaires as soon as possible. If parents did not respond, reminders were sent out (see procedure above).

Finally, families also received 'Thank-you' gifts for participation. After questionnaire completion, parents were sent a personalised word-cloud via post at 12, 18 and 24 months (see anonymised example in Appendix 6). After completing the in-person sessions, participants received a Babylab T-shirt at 18 months and a £5 shopping voucher at 24 months as reimbursement for time and expenses. At 36 months, families received a personalised participation certificate and a £5 shopping voucher.

5.1.4. Ethics

Ethical approval was obtained by the University of Lincoln's School of Psychology Research Ethics Committee. Parents' details were recorded in a database which was kept confidential, stored in a locked cabinet and only accessible to Babylab researchers. The researcher explained to each participant that their participation was voluntary and that they could cease taking part at any point during the longitudinal research and that they could withdraw their data until before publication. Participants' data was anonymised, and no personal information was ever disclosed in any publications. An amendment to the Ethics proposal for taking videos of the child and caregiver during the in-person session was submitted on 16th June 2015 and approved. Parents were assured that the videos would be stored securely in a locked filing cabinet at the University or on University secure servers and not shared with others outside the Lincoln Babylab (see Appendix 7 for consent form).

5.1.5. Statistical analysis

The results wil be presented in four parts, along with the respective research questions. In the first part, the descriptive data is presented as first step. Then the normality of the data is checked using Kolmogorov-Smirnov tests of normality, and data transformation is conducted where required (two-step transformation, see Templeton, 2011). For analyses with missing data cases were removed for those analyses but participants' data was included in other analyses in which they had no missing data. It is further investigated if differences between boys and girls can be found in terms of language scores using ANCOVAs to control for the age at testing.

The relationships between UK-CDI scores and later language scores is investigated as a next step. Due to the high variation in age at Time 1 (around 12 months) partial correlations are used for correlations between UK-CDI scores at Time 1 and later language scores to control for age at Time 1. For the correlations between UK-CDI scores at Time 2 (18 months) and later language Pearson correlations are used.

In analysis 2, MANOVAs (multivariate analysis of variance) were employed to assess if communication ability level had a significant effect on earlier, concurrent or later language scores. Analysis 3 assessed how much variance UK-CDI scores (i.e. Comprehension, Production, Gestures and Phrases Understood) could contribute to later language scores using hierarchical multiple regressions. In addition, separate hierarchical regressions were used to establish the contributions of the predictor variables (i.e. prematurity, gender, SES, sibling status, ear infections, sleep, family risk status of dyslexia or speech or language problems, cognitive and motor skills) to later language skills after controlling for early language ability. For some more in-depth analyses, the influence of the different family, biological or environmental factors on language development was further assessed using MANOVAs/ MANCOVAs, Spearman correlations and moderation analyses depending on the type of data. Analysis 4 used receiver operating characteristic curves (ROC) to establish if UK-CDI scores at 12 and 18 months could be used to predict language delay at 24 and 36 months. For Parts 2 - 4, more details about statistical analyses are given in respective chapters.

5.2. RESULTS: Descriptive Statistics, distribution of the data and gender differences between communication scores

Most studies investigating the predictive validity of CDIs reported the relationship between CDI scores and later language scores. To compare such previous research with this longitudinal study, correlations between the different measures are described from 12 months (Time 1) to 36 months (Time 4).

5.2.1. Language assessed with UK-CDI for infants, Lincoln Toddler CDI and 3-year parent report language measure

Table 19 shows the descriptive statistics for the different categories on the UK-CDI at 12 months (Time 1) and 18 months (Time 2) and the vocabulary and grammar scores for the Lincoln Toddler CDI at 24 months (Time 3) and the 3-year parent report language measure at 36 months (Time 4). When comparing the data in Table 19 below, it needs to be considered that the questionnaires use different numbers of items to measure concepts (e.g. comprehension, production and grammar) and therefore percentage scores should be used for comparison.

Table 19. *Descriptive Statistics for UK-CDI scores at 12 months, 18 months and vocabulary and grammar scores at 24 and 36 months, separated by gender*

		Total					Boys					Girls				
Age (Time point)	Communication variable	N	M (SD)	Median	Range	M%	N	M(SD)	Median	Range	M%	N	M(SD)	Median	Range	M%
	Gestures															
12 months (Time 1)	UK-CDI	82	23.30 (8.59)	22.5	9.5-48.5	**36.98%**	49	22.47 (8.82)	21.5	11-48.5	**35.67%**	33	24.53 (8.23)	24	9.5-45	**38.94%**
18 months (Time 2)	UK-CDI	80	44.13 (8.48)	43.75	20-62	**70.05%**	47	42.01 (8.48)	42.5	20-60	**66.68%**	33	47.14 (7.64)	47	30-62	**74.83%**
	First Signs of Understanding															
12 months (Time 1)	UK-CDI	82	2.98 (.16)	3	2-3	**99%**	49	2.98 (.14)	3	2-3	**99%**	33	2.97 (.17)	3	2-3	**99%**
18 months (Time 2)	UK-CDI	82	3 (0)	3	3-3	**100%**	49	3 (0)	3	3-3	**100%**	33	3 (0)	3	3-3	**100%**
	Phrases Understood															
12 months (Time 1)	UK-CDI	82	11.23 (6.11)	10.5	1-27	**40.11%**	49	11.29 (6.36)	10	1-27	**40.32%**	33	11.15 (5.81)	11	1-25	**39.82%**
18 months (Time 2)	UK-CDI	82	22.18 (4.77)	23	11-28	**79.21%**	49	21.43 (5.19)	23	11-28	**76.54%**	33	23.30 (3.89)	24	16-28	**83.21%**
	Comprehension															
12 months (Time 1)	UK-CDI	82	67.72 (58.85)	49	0-313	**17.14%**	49	70.49 (66.03)	46	0-313	**17.85%**	33	63.61 (46.87)	55	6-203	**16.10%**
18 months (Time 2)	UK-CDI	82	211.23 (86.06)	210.5	29-395	**53.48%**	49	206.12 (96.12)	216	29-395	**52.18%**	33	218.82 (69.19)	209	92-362	**55.40%**
24 months (Time 3)	Lincoln Toddler CDI	82	440.48 (128.23)	441	157-682	**63.93%**	49	429.51 (140.78)	402	157-679	**62.34%**	33	456.76 (106.91)	459	253-682	**66.29%**
	Production															
12 months (Time 1)	UK-CDI	82	8.35 (14.43)	4	0-99	**2.11%**	49	8.96 (16.89)	5	0-99	**2.27%**	33	7.45 (9.91)	4	0-54	**1.89%**
18 months (Time 2)	UK-CDI	82	63.09 (64.78)	39.5	2-316	**15.97%**	49	58.86 (72.61)	32	2-316	**14.90%**	33	69.36 (51.44)	60	2-189	**17.56%**
24 months (Time 3)	Lincoln Toddler CDI	82	297.30 (169.39)	298	20-676	**43.15%**	49	286.96 (172.41)	297	20-617	**41.65%**	33	312.67 (166.22)	299	24-676	**45.38%**
36 months (Time 4)	3-year parent report language measure	82	71.66 (26.09)	82	0-100	**71.66%**	49	68.51 (28.10)	78	0-100	**68.51%**	33	76.33 (22.38)	83	21-100	**76.33%**
	Sentence Complexity															
24 months (Time 3)	Lincoln Toddler CDI	82	8.61 (9.75)	4	0-33	**23.27%**	49	7.45 (9.52)	2	0-33	**20.14%**	33	10.33 (9.97)	7	0-30	**27.92%**
36 months (Time 4)	3-year parent report language measure	82	8.05 (3.99)	9	0-12	**67.08%**	49	7.45 (4.45)	9	0-12	**62.08%**	33	8.94 (3.03)	10	2-12	**74.50%**
	Advanced Grammar															
36 months (Time 4)	3-year parent report language measure	82	8.52 (3.10)	9	0-12	**71.00%**	49	8.06 (3.53)	9	0-12	**67.17%**	33	9.21 (2.19)	10	2-12	**76.75%**

At 12 months, most children were able to master the First Signs of Understanding questions. Due to ceiling effects we do not make use of this subscale for analysing the predictive validity of the UK-CDI. The data also shows that children's knowledge increases for all other subscales over time. The ability to combine words is an important indicator of early grammar and this forms one question in the questionnaires used at 24 and 36 months (Has your child begun to combine words yet, such as "nother cracker", or "doggie bite"?). The answers are not listed in Table 19 but will be described here. At 24 months, more than half the children often combined words in spontaneous speech (60.98%), some children used word combinations sometimes (26.83%) and the rest were not yet combining words (12.20%) according to the Lincoln Toddler CDI. One year later, most children usually used word combinations in their typical communication except for two children who only used them sometimes (2.4%) – this was measured using the 3-year parent report language measure.

5.2.1.1. Data distribution and transformation

Gesture scores were normally distributed at 12 and 18 months. For First signs of understanding there was little variance in the scores at 12 months, and the scores reached ceiling at 18 months. Scores for Phrases Understood were not normally distributed at 12 and 18 months according to the Kolmogorov-Smirnov test of normality (see overview of normality test for all language variables in Appendix 8). The scores were positively skewed (left-hand side of the graph) at 12 months and negatively skewed at 18 months (right-hand side of the graph).

Comprehension scores were not normally distributed at 12 months, but were normally distributed at 18 and 24 months according the Kolmogorov-Smirnov test of normality. Comprehension scores at 12 months were positively skewed and showed positive kurtosis indicating a peaked distribution with long thin tails to the right of the graph. There were also two outliers (+ 2.5 SD from the mean). The distribution was still not normal after removing the outliers.

For Production scores, the Kolmogorov-Smirnov test of normality was always significant indicating no normal distribution between 12 and 36 months. The Production scores showed a strong positive skew and positive kurtosis at 12 months and less so at 18 months. There were three outliers at 12 months and one outlier at 18 months (+ 2.5 SD from the mean). The distribution was still not normal after removing the outliers. From the Histogram, the production scores at 24 months resembled a normal distribution most. At 36 months, production scores were negatively skewed and there were no outliers.

The grammar scores were positively skewed at 24 months and negatively skewed at 36 months according to the Kolmogorov-Smirnov test and there were no outliers.

For the following analyses all language scores were transformed into z-scores and the skewed data underwent two-step transformation (Templeton, 2011). This technique involves the transformation into z-scores as a first step, and the normalisation of the data in a second step. This transformation technique was chosen as log and square root transformation alone could not normalise most of the data. It should also be mentioned here that some authors think that skewed data should not be transformed in general (for a discussion see Field, 2017, pages 268 - 270). For our data, transforming the data was successful with few exceptions (PLS Expressive Communication at 18 months, Sentence Complexity at 24 and 36 months, ASQ Communication at 36 months).

5.2.1.2. Gender differences

We analysed if there were significant differences between boys and girls for any of the communication measures at any time point. ANCOVAs of gender on the various measures (see Table 19) with age as covariate were used to control for the effects of age. Gender was the independent variable (IV) and the dependent variables (DV) were the different CDI-type measures. Preliminary checks were conducted to ensure that there were no violations of ANCOVA assumptions (for more information about assumptions of ANCOVA, see Pallant, 2013, pages 309 - 311). The assumption of

Levene's test of equality of error variances was only violated for comprehension at 18 months and 24 months and grammar scores at 36 months, thus independent t-tests were conducted for these variables.

Only Gesturs showed a significant difference between boys and girls with girls using significantly more Gestures at 18 months than boys, but the effect size was small, $F(1,77) = 7.67$, $p = .007$, partial eta squared = .03. No other significant gender differences were found for any of the communication scores, see Appendix 9.

The data was analysed for outliers separately for boys and girls and it was found that there were no outliers in either gender group. Therefore, the reason for small gender differences was not due to increased mean scores by outliers at the top of the distribution for boys.

5.2.2. Preschool Language Scale (PLS-5 UK)

At 18 months and 24 months, the PLS-5-UK was used as an independent tool to assess children's Auditory Comprehension, Expressive Communication and Total Language. For the descriptive statistics see Table 20 below.

Table 20. *Descriptive Statistics for PLS scores at 18 and 24 months, separated by gender*

	Total					Boys					Girls				
Communication variable	N	M (SD)	Median	Range	M%	N	M (SD)	Median	Range	M%	N	M (SD)	Median	Range	M%
Auditory Comprehension															
18 months (Time 2)	79	105.65 (14.81)	107	73-147	**70.43%**	48	103.31 (15.42)	105	73-133	**68.87%**	31	109.26 (13.25)	110	84-147	**72.84%**
24 months (Time 3)	78	103.22 (11.83)	103	66-139	**68.81%**	47	100.62 (12.40)	100	66-124	**67.08%**	31	107.16 (9.82)	106	89-139	**71.44%**
Expressive Communication															
18 months (Time 2)	79	101.86 (8.17)	98	91-129	**67.91%**	48	100.88 (7.39)	98	91-121	**67.25%**	31	103.39 (9.15)	102	91-129	**68.92%**
24 months (Time 3)	78	101.27 (9.66)	103	85-133	**67.51%**	47	99.62 (8.87)	100	85-113	**66.41%**	31	103.77 (10.39)	103	85-133	**69.18%**
Total Language															
18 months (Time 2)	79	104.04 (11.22)	104	81-137	**69.36%**	48	102.21 (11.19)	103.5	81-128	**68.14%**	31	106.87 (10.84)	105	88-137	**45.38%**
24 months (Time 3)	78	102.47 (10.42)	103	79-139	**68.31%**	47	100.17 (10.39)	102	79-117	**66.78%**	31	105.97 (9.59)	105	89-139	**70.64%**

5.2.2.1. Distribution

For Auditory Comprehension, there were eight children who scored 1 SD below the mean (i.e. score range 73 - 84) at 18 months and at 24 months two children scored between 1 - 2 SD below the mean (i.e. scores of 82) and two children scored below 2 SD (i.e. score range 66 - 69). For Expressive Communication, no children scored more than 1 SD below the mean at 18 months. At 24 months, there were nine children who scored exactly 1 SD below the mean (i.e. score of 85). There was one outlier at the top of the distribution; no other PLS-5 UK data had outliers. For the Expressive Communication at 18 months, there was one outlier value (129) which was changed to the next lowest (non-outlier) number (121) (Pallant, 2013; Tabachnick & Fidell, 2001). This data was also positively skewed which did not change after the outlier was modified; the other PLS scores were all normally distributed or close to being normally distributed (Auditory Comprehension at 24 months).

For Total Language, there were two children who scored between 1 - 2 SD and one child who scored exactly 1 SD below the mean at 18 months. Five children scored between 1 - 2 SD below the mean at 24 months.

5.2.2.2. Gender differences

Girls had higher mean scores than boys for the PLS-5 at 18 and 24 months, see Table 20. Independent-samples t-tests (IV= gender, DV = Auditory Comprehension at 24 months / 18 months) showed a significant difference for Auditory Comprehension at 24 months (t (76) = -2.47, p = .016, two-tailed), but not at 18 months, and the magnitude of the differences in the means (mean difference = -6.54, 95% *CI*: -11.82 to -1.27) was moderate (eta squared = 0.07). Independent-samples t-tests for gender (IV) and ExpressiveCommunication at 24 months (DV) showed that the difference in the means approached significance (t (76) = -1.89, p = .062, two-tailed) and the effect was small (mean difference = -4.16, 95% *CI*: -8.53 to .22, eta squared = .04). The independent-samples t-test between gender (IV) and Total Language at 24 months (DV) found a significant difference (t (76) = -2.49, p = .015, two-tailed) and the effect was moderate (mean difference = -5.80, 95% *CI*: -10.44 to -1.15, eta squared = .08), see Appendix 10.

5.2.3. Language assessed with Ages and Stages Questionnaire (ASQ)

The ASQ was used only at 36 months (Time 4) as an independent tool to measure children's language at this higher age and to distinguish between delayed and non-delayed children. The communication subscale data is presented below in Table 21.

5.2.3.1. Distribution

The data is naturally negatively skewed which means that a distinction between skilled and very skilled communicators cannot be made. There were no outliers in the data. However, four children scored 2 SD below the mean and six children scored between 1 - 2 SD below the mean. Out of those 10 children, all scored well below 1 SD below the mean on the vocabulary scale of the 3-year parent report language measure but out of the 4 delayed children two scored within in the normal range (above 1 SD) in one of the grammar tasks (Sentence Complexity or Advanced Grammar) at 36 months. On the other hand, there were 4 children who scored 2 SD below the mean on the vocabulary measure of which the ASQ recognised two also as delayed, one as mildly delayed and one within in the typical range. This typical child also scored within the typical range on the Advanced Grammar scale. This shows that a range of different skills need to be considered at this age before making decisions about language status.

Table 21. *Descriptive Statistics for ASQ at 36 months, separated by gender*

	Total					Boys					Girls				
Communication variable	N	M (SD)	Median	Range	M%	N	M (SD)	Median	Range	M%	N	M (SD)	Median	Range	M%
ASQ 36 months (Time 4)	82	52.87 (8.57)	55	20-60	**88.12%**	49	51.94 (9.40)	55	20-60	**86.57%**	33	54.24 (7.08)	55	30-60	**90.40%**

5.2.3.2. Gender differences

Differences between boys and girls were analysed. An independent t-test with gender (IV) and ASQ scores at 36 months (DV) was conducted rather than an ANCOVA as the ANCOVA assumption of homogeneity of regression slopes was violated. The results showed that girls (M = 1.67, SD = .06) and boys (M = 1.68, SD = .08, t (80) = 1.17, p = .247, two-tailed) did not differ significantly in terms of ASQ communication scores at 36 months and the effect was small (mean difference = .01, 95% CI: -.01 to .05, eta squared = .02), see Appendix 11.

5.2.4. Correlations between early UK-CDI scores and later language measures

Correlations between the UK-CDI scores[40] (at 12 and 18 months: Gestures, Phrases Understood, Comprehension and Production) with prospective communication and language scores were calculated between 12, 18, 24 and 36 months using the above described range of language measures at follow-up (UK-CDI, Lincoln Toddler CDI, 3-year parent report language measure, PLS-5UK, ASQ). Partial correlations were used in the following to control for the effect of age at 12 months (covariate) for correlations which included UK-CDI variables at 12 months. According to the Bonferroni correction, the alpha level was set to .003 to correct for multiple comparisons because there were 17 correlations per variable (0.05/17 = .003). However, as the Bonferroni correction is very conservative, I focus more on the strength of the effect sizes when discussing the results. Please see Table 22 for correlations.

[40] All UK-CDI scores were used except for the First signs of understanding subscale as most children (99%) passed this subscale at 12 and 18 months and due to these ceiling effects correlations would not give meaningful results. Other research studies have not used this subscale for assessing the predictive validity of CDIs either (see literature review in the introduction).

Table 22. *Correlations between UK-CDI scores at 12 months (partial correlations: contro lling for age at 12 months) and 18 months and later communication and language variables*

UK-CDI categories		UK-CDI Phrases Understood 18m	UK-CDI Gestures 18m	UK-CDI Comprehension 18m	UK-CDI Production 18m	PLS Auditory Comprehension 18m	PLS Expressive Communication 18m	PLS Total language 18m	Lincoln TCDI Comprehension 24m	Lincoln TCDI Production 24m	Lincoln TCDI Sentence Complexity 24m	PLS Auditory Comprehension 24m	PLS Expressive Communication 24m	PLS Total language 24m	3-year parent report measure Production 36m	3-year parent report measure Sentence Complexity 36m	3-year parent report measure Advanced grammar 36m	ASQ communication 36m
Gestures 12m	Correlation	.41***	.61***	.47***	.33**	.22	.28*	.27*	.34**	.25*	.31**	.21	.11	.18	.14	.15	.17	.14
Gestures 18m	Correlation								.36***	.32**	.33**	.40***	.20	.35**	.21	.15	.26*	.26*
Phrases Understood 12m	Correlation	.55***	.38***	.60***	.34**	.24*	.31**	.30*	.43***	.33**	.35**	.13	.19	.18	.26*	.27*	.39***	.24*
Phrases Understood 18m	Correlation								.52***	.43***	.44***	.45***	.30**	.43***	.42***	.37***	.37***	.39***
Comprehension 12m	Correlation	.39***	.33**	.58***	.30*	.26*	.29*	.30*	.40***	.28*	.36**	.12	.21	.17	.19	.21	.24*	.11
Comprehension 18m	Correlation								.77***	.66***	.57***	.50***	.49***	.56***	.54***	.38***	.45***	.39***
Production 12m	Correlation	.35**	.18	.41***	.50***	.11	.33**	.20	.42***	.32**	.47***	.08	.19	.15	.26*	.35***	.31**	.21
Production 18m	Correlation								.65***	.69***	.72***	.35**	.58***	.51***	.49***	.47***	.47***	.44***

a. Cells contain zero-order (Pearson) correlations.

Significance is 2-tailed

*p < .05, **p < .01, ***p < .001

5.2.4.1. Gesture correlations

The strongest correlation for Gestures at 12 months was found with Gestures at 18 months, this association was significant and large, see Table 22. Gestures at 12 months correlated moderately with the other UK-CDI scores at 18 months but the associations were only weak with the PLS-5 UK at the same age. The strength of correlation decreased over time as most associations between Gestures at 12 months with language scores at 24 and 36 months were weak except for the Lincoln Toddler CDI measures of Comprehension and Sentence Complexity at 24 months which showed moderate correlations with Gestures at 12 months.

When Gestures were assessed at 18 months, they correlated moderately with most language scores six months later except for the PLS Expressive Communication scale for which weak associations were found with Gestures at 18 months. As seen above, the correlations decreased over time and at 36 months several variables showed weak correlations with Gestures at 18 months.

Overall this shows, that stronger correlations were found between the same domain (Gestures at 12 and 18 months), if the time gap was shorter compared to longer and with other CDI scores compared to in-person assessments (particularly for the Expressive Communication subscale of the PLS-5 UK rather than the Auditory Comprehension subscale). In sum, the correlational reach for Gestures at 12 months went just into 2 years, but the reach of Gestures at 18 months went right up til 36 months.

5.2.4.2. Phrases Understood

Phrases Understood at 12 months showed strong correlations with Phrases Understood and Comprehension at 18 months and moderate correlations with the other UK-CDI scores at 18 months. The correlations between Phrases Understood at 12 months and PLS-5 UK scores at 18 months were somewhat weaker showing weak to moderate associations. Phrases Understood at 12 months also correlated significantly and moderately with all CDI scores at 24 months but weakly with PLS scores at that age. In addition, Phrases Understood at 12 months was significantly and moderately associated with grammar (Advanced Grammar) at 36 months but all other correlations were weak at this age.

At 18 months, Phrases Understood showed stronger and more consistent associations with all later language scores than at 12 months. Phrases Understood at 18 months correlated significantly and generally moderately with all later scores except for Comprehension at 24 months which showed a large correlation with Phrases Understood at 18 months.

Overall, the correlations were stronger with the CDI than with in-person scores across all situations. Stronger correlations were found with Comprehension than with Production, this could be because Phrases Understood is more related to the Comprehension category than to Production.

5.2.4.3. Comprehension correlations

Comprehension at 12 months correlated significantly and strongly with Comprehension scores at 18 months. Other correlations between Comprehension at 12 months and CDI scores at 18 months were weak to moderate (Production, Gestures, Phrases Understood at 18 months). Furthermore, associations between Comprehension at 12 months and PLS scores at 18 months existed, but were weak. Comprehension at 12 months also correlated moderately with Comprehension and Sentence Complexity at 24 months. Other correlations for Comprehension at 12 months with the other CDI or PLS scores at 24 months were weak. In addition, there were associations between Comprehension at 12 months and all language scores at 36 months, albeit weak ones.

Interestingly, Comprehension scores at 18 months showed strong correlations with all CDI and PLS scores at 24 months, except for the Expressive Communication scale of the PLS at 24 months (moderate effects). Importantly, correlations between Comprehension at 18 months and

Production at 36 months was also strong. In addition, there were moderate effects with grammar (3-year parent report) and general language (ASQ) at this age.

In summary, the relationship between Comprehension scores at 18 months was strong with later Production scores but Comprehension scores at 12 months correlated weakly with later Production scores. This indicates that the age of testing language ability for predicting later language is very important. Again, the correlations were somewhat stronger for the CDI measures compared to the PLS assessments particularly for Comprehension at 12 months.

Moving on to 18 months, it is visible that the correlation between Comprehension at 18 months and PLS Auditory Comprehension scores (and PLS total scores) at 24 months were also strong. In addition, the relationships were stronger if the testing happened close to the first testing and decreased over time in strengths. For example, the strongest correlation was between Comprehension at 18 and 24 months, as Comprehension at 24 months helped to explain 58% of the variance in the Comprehension scores six months later. It should also be mentioned that whilst the strengths of association decreased over time, Comprehension at 18 months still correlated significantly and moderately with all language scores (3-year parent report language measure and ASQ) at 36 months.

In sum, early Comprehension (12 months) had a moderate correlational reach up until 2 years, except for not being correlated to the PLS. And later Comprehension (18 months) had initially a strong reach with scores at 24 months, and then still a strong to moderate reach for all following tests including at 36 months.

5.2.4.4. Production correlations

Production at 12 months was significantly and moderately correlated with the UK-CDI scores at 18 months except for Gestures at 18 months. Furthermore, Production at 12 months moderately correlated with the PLS Expressive Communication scores at 18 months, whilst the relationship with the other PLS scores was weak at this age. This is because Production which measures expressive vocabulary and PLS Expressive Communication which also measures expressive vocabulary as well as other means of expressive communication (e.g. gesture and sound production) are more closely related than Production with PLS Auditory Comprehension. The relationship between Production scores at 12 months and CDI scores at 24 months was moderate. Again, the association was weaker for the PLS scores, as Production at 12 months correlated weakly with PLS scores at 24 months. Furthermore, Production at 12 months and grammar (Sentence Complexity and Advanced Grammar) at 36 months correlated moderately whilst Production at 12 months was only weakly associated with Production and a measure of general language ability (ASQ) at 36 months.

Interestingly, Production at 18 months correlated significantly and strongly with all CDI and PLS scores at 24 months, except for one, Auditory Comprehension for which the association was of moderate strengths. Furthermore, Production at 18 months also correlated moderately with Production and grammar (3-year parent report) as well as general language (ASQ) at 36 months.

In summary, in comparison to the previous categories the strengths of relationships stayed stronger over time for Production at 12 and at 18 months. In addition, stronger correlations were found with CDI scores compared to the in-person measures of which Expressive Communication generally showed stronger correlations than Auditory Comprehension (see explanation above).

Overall, early Producton (12 months) had a moderate correlational reach up until 2 years, except for not being correlated to the PLS, and a weak to moderate correlational reach until 36 months with the 3-year parent measure. Later Production (18 months) had a strong reach with scores at 24 months, and then still a moderate correlational reach for all following tests at 36 months.

5.3. DISCUSSION

The short-term and long-term associations were reported for the main four subscales of the UK-CDI. Overall word knowledge (Comprehension, Production and Phrases Understood) at 12 months correlated with later language but stronger correlations were found for word knowledge at 18 months with later language scores.

Interestingly, UK-CDI word knowledge at 18 months (Comprehension, Production, Phrases Understood) showed highly significant correlations than knowledge at 12 months with later word and language measures at 24 and 36 months. Other research found similar results (e.g., Unhjem et al., 2015).In contrast, Gesture scores were also related to later language but with less strength as children got older. Bavin et al. (2008) found similar results.

Generally, stronger associations were found if the same linguistic category was tested at follow-up. However, there were also few exceptions, for example Comprehension at 18 months was associated strongly with Production at 36 months whilst Production at 18 months was only moderately correlated to Production at 36 months. Other research also suggested that early Comprehension is better than early Production at predicting later Production scores (e.g., Thal et al., 2013). Furthermore, UK-CDI scores correlated better with Toddler CDI questionnaire data rather than in-person testing scores using the PLS-5 UK, this difference was also found by other authors (Guiberson, 2008; P. Lyytinen et al., 1996).

5.3.1. Comprehension

The relationships between Comprehension scores at 12 and 18 as well as between 18 and 24 months were large in strengths and between 12 and 24 months they were moderate. This is supported by other studies which looked at similar age groups (Bornstein et al., 2006; Hamadani et al., 2010; Unhjem et al., 2015).

Our study found weak correlations between Comprehension at 12 months and later Production scores and moderate to weak relationships between Comprehension at 12 months and later grammar scores. This is in line with several other studies that found similar results (P. Lyytinen et al., 1996; P. Lyytinen et al., 1999; Rose et al., 2009; Sachse, Saracino et al., 2007; Unhjem et al.,

2015) with few exceptions where Comprehension and later Production correlated moderately in studies with larger samples (Bavin et al., 2008; Feldman et al., 2000).

However, Comprehension tested at 18 months correlated strongly with later Production at 24 and 36 months and strongly to moderately strong with grammar. This result was also found by other studies (Cochet & Byrne, 2016; Friend et al., 2012). In contrast, Production at 18 months correlated only moderately with Production at 36 months. This result is in line with previous research with suggests that Comprehension was somewhat better than Production at predicting later Production scores (e.g., Thal et al., 2013).

Whilst the relationships between Comprehension at 12 months with PLS scores at 18 or 24 months were weak, the association between Comprehension at 18 months and especially Auditory Comprehension at 24 months was strong. This is an interesting finding as it shows that Comprehension (which measures receptive vocabulary) is more stable from 18 months onwards compared to 12 months. Furthermore, the PLS assessment measures more precursor abilities to later language comprehension at 18 months (i.e. play behaviour, understanding of gestures) compared to at 24 months when language comprehension can be measured more directly (i.e. measuring word and language comprehension from questions using syntactically advanced constructions). This suggests that the concepts are more similar for Comprehension at 18 months and PLS Auditory Comprehension at 24 months which could be another reason for stronger associations between these ages.

5.3.2. Production

Production at around 12 months was moderately related to later Comprehension, Production and grammar scores. The results are in line with previous research which looked at early Production and later Comprehension (Korpilahti et al., 2016; Unhjem et al., 2015), later Production (Bavin et al., 2008; Bornstein et al., 2006; Feldman et al., 2000; Hamadani et al., 2010; Korpilahti et al., 2016; Rose et al., 2009; Sachse, Saracino et al., 2007) or later grammar (Feldman et al., 2000; Kreisman, 2012). Smaller correlations were expected for Production at 12 months with later language scores compared to other communication variables at 12 months with later language. Our results support this assumption, see Table 22. This is because at 12 months there were still floor effects for Production as children only started to say their first words at that age. According to the literature (Thal et al., 2013), most children with an initial Production delay caught up with their peers at kindergarten age which is how the decreasing correlations with Production at 36 months may be explained. However, these delayed children usually developed grammar more slowly, thus a stronger relationship between Production and later grammar was expected which was confirmed by our study.

Furthermore, Production at 18 months showed mostly strong relationships with the different language variables at 24 months and moderate relationships with language variables at 36 months. Similar results were found in other studies which looked at relationships between Production scores during the second year and later Production (P. Lyytinen et al., 1999; Thal et al., 2013; Unhjem et al., 2015) as well as later grammar scores (Thal et al., 2013). At 18 months, all children in our sample had begun to produce words but the range of ability between children was large. Furthermore, at this age language started to become more stable as all relationships even with in-person assessments showed significant correlations with moderate to large effect sizes.

5.3.3. Gestures

The Gestures and Phrases Understood subscales are usually only used during infancy, whilst vocabulary (Comprehension and Production) can be assessed through to adulthood using standardised assessments (Brownell, 2000a; Brownell, 2000b; Williams, 2007). The strongest correlations were usually found between the same categories according to the literature review (see introduction page 105), hence smaller associations were expected for Gestures and Phrases Understood with language variables at 24 and 36 months compared to Production and Comprehension.

The results showed that Gestures correlated moderately with later Gestures scores confirming the results of several other studies (Bavin et al., 2008; Fenson et al., 1994; Fenson et al., 2007). In our study, Gestures were also moderately correlated with Comprehension and Auditory Comprehension scores which were studied up to 24 months. Somewhat lower associations were found between symbolic gestures (selected items from Gestures scale) between 14 and 18 months using an in-person assessment (Laakso et al., 1999; P. Lyytinen et al., 1996). Symbolic gestures are later acquired gestures. This means that at 14 months there might not yet have been a lot of variability between children, explaining weaker associations with later language variables. Furthermore, our data also showed weak correlations between Gestures (at 12 months) and in-person assessments (PLS-5 UK) at 18 months compared to moderate to strong relationships with UK-CDI variables at the same age. Weak correlations were recorded between Gestures and Comprehension if tested between 12 and 36 months (Kreisman, 2012; Rose et al., 2009) and no significant associations were found at older ages (Fish & Pinkerman, 2003). Our correlations between 12 and 18 months were also stronger than between 12 and 24 months. Futhermore, some weak associations between Gestures at 18 months and grammar as well as general language (ASQ) at 36 months were found. This supports the notion that the strengths of associations decrease over time. This is because most of the initially delayed children caught up with their peers at kindergarten age particularly in terms of Production and Comprehension (Thal et al., 2013). Gestures also correlated significantly and moderately with Production between 12 and 18 months and with Production and Sentence Complexity between 18 and 24 months as found by Thal et al. (2013). In contrast to some other studies conducted in Australia, USA and Germany who reported moderate relationships (Bavin et al., 2008; Feldman et al., 2000; Sachse, Saracino et al., 2007), we found a weak correlation between Gestures and Production scores between 12 and 24 months. Furthermore, Gestures were only weakly correlated with language variables at around 3 years. Similar results were found by Rose et al. (2009). Other research has found that Gestures at 15 months were not correlated with risk of language impairment at kindergarten age; however, a Structural Equation Model could show that Gestures had an indirect effect as the influence of

Gestures on language impairment risk was mediated by Production at 15 months (Hsu & Iyer, 2016). This indicates that early Gestures should be taken into consideration when aiming to predict language delay in early childhood.

5.3.4. Phrases Understood

When comparing the results of Phrases Understood with the results by Feldman et al. (2000), it was found that the correlations were very similar for Production (.33 vs .31) and somewhat stronger for Sentence Complexity in our study (.35 vs .18) which showed a significant, moderate relationship in comparison to a weak relationship by Feldman (2000). It should be noted that Feldman (2000) looked at a more racially and SES diverse group than used in our sample (see explanation above for Kreisman (2012)). To my knowledge, there is only one more study which looked at Phrases Understood but they tested children at 15 months and at 4 and 5;4 years (Fish & Pinkerman, 2003). Significant and weak relationships were found between Phrases Understood at 15 months and Expressive and Total Language as measured by the PLS-3 at 4 years. This result cannot be directly compared to our findings due to the ages of testing; however, our data indicate moderate relationships between Phrases Understood at 12 and 18 months with language at 36 months thus links with later language seem possible but would need to be tested to be confirmed. Taken together, these results indicate that the early ability of Phrases Understood shows some continuity in terms of later language skills. Therefore, I will include this subscale for establishing the predictive validity of the UK-CDI where possible.

5.3.5. Summary

Three main results stand out: The correlations were usually strongest a) with the same category b) at the closest follow-up testing (all significant and strong); c) correlations weaken with increasing age gap between testing times. Other studies (Bavin et al., 2008; Feldman et al., 2000; P. Lyytinen et al., 1996) came to the same conclusion. In addition, the strengths of associations within the same category generally decreased over time. This also supports the results by Unhjem et al. (2015) who found a decrease in the strength of association for the same category over time for the control group representing typically developing children in contrast to the at-risk group who maintained high associations over two testing intervals (each 3 months apart).

In addition, UK-CDI scores taken at around 12 months showed stronger associations up to 24 months compared to 36 months. However, it should be noted that Phrases Understood and Production at 12 and at 18 months correlated significantly and moderately with grammar scores at 36 months - in other studies this was only assessed for Production up to 36 months, and similar results were obtained (Thal et al., 2013).

The data also show that language was more stable from 18 months onwards as the correlations were consistently significant with scores at 24 and 36 months. Other research has also shown that correlations were stronger when children were older at the first time of testing (Cochet & Byrne, 2016; Friend et al., 2012).

Taken together, results of the newly established UK-CDI largely support the findings of other CDI predictive validity studies. This is the first time the UK-CDI has been used to measure correlations with later language skills, and it is important as the UK-CDI is the only comprehensive communicative inventory which has been normed and standardised for the British population. In addition, this study adds value to the research field as it includes all relevant CDI: Words and Gestures scales (including Phrases Understood) and regular testing points using a variety of parent report and in-person assessments whilst other studies often used fewer testing points and a smaller variety of testing materials.

The data suggests a strong continuity between time points particularly between the UK-CDI scores at 18 months and later language scores. For this reason, I will also examine if these UK-CDI scores can be used to make predictions about later language delay for individual children.

6. ANALYSIS 2: ABILITY GROUPS AND LANGUAGE STABILITY OVER TIME

Further analysis was conducted to investigate the current UK-CDI data in terms of ability groups and language stability over time. Research has shown that the prevalence of children between 2 and 3 years with significant language delays not due to underlying hearing impairments, neurological or emotional disorders or genetic syndromes is between 15% - 20% (Horwitz et al., 2003; Reilly et al., 2007). At 5 years about 6.8% of children still have a language deficit (for a review see Law, Boyle, Harris, Harkness, & Nye, 2000b, see also Introduction, pages 5 - 6). This decrease of language deficits over time has to be taken into account when aiming to detect children with later language delay/disorder during early development. During the early stages of language development (around 12 months), there is little variance in language production scores. However, shortly after that period and throughout toddlerhood language ability is characterized by a high variability between children.

Due to this smaller variance in early language, children were grouped into quartiles (at 12 and 18 months) to distinguish them by ability. This choice of cut-off was adopted following criteria by a recent study by Duff et al. (2015) who also classified British English children aged 18-19 months as late-talkers if they fell below the 25th percentile on the Oxford CDI. Thus, for this next analysis, I will use quartile groups which may show a clearer developmental picture and stronger stability over time than looking at small groups at the extreme ends of the spectrum only who show some stability, with suggestions that high language ability (at or above 90th percentile) at early stages is more stable over time (Thal et al., 1997). It is important to follow up on Duff's approach to investigate if language is also stable if children are grouped into wider groups (e.g. at or below the 25th percentile rather than at or below the 10th percentile) and for groups with average ability to include a larger amount of children overall.

Given previous research and the results presented in Chapter 5, Analysis 1, above, the hypotheses were as follows:

1) Children from the higher language ability groups continue to have higher language scores at later timepoints and that children from medium and lower ability groups continue to have lower language scores than high ability children.
2) High language ability level should be more stable over time than the other ability groups. This means children with high abilities during early communication development are more likely to remain advanced in their language skills when tested at older ages. In contrast, children with typical or slow initial language show less continuity over time and thus are more likely to change group membership.
3) More stability was expected within the same language categories compared to between categories (e.g. Production ability was expected to have a significant effect on later Production scores).
4) Better prediction of scores was expected using Comprehension and Production than Gesture ability groups.
5) More stability was expected from ability groups at 18 months as opposed to ability groups at 12 months.

6.1. METHOD

6.1.1. Participants

Participants with missing values (7%) were omitted so that the data set included 67 children[41]. The children (28 girls) were aged 11.58 months on average at Time 1 (SD = 1.06, range = 10 - 15 months). Levels of ability were created from the UK-CDI preliminary norms (Alcock et al., 2017) for Gestures, Production and Comprehension at 12 and 18 months. For example, the preliminary norms

[41] In total 15 children were not taken into account for this analysis. Ten children were 19 months at Time 2 but UK-CDI (preliminary) norms are for children between 8-18 months. Due to lack of reference data, these children could not be put into ability groups. In addition, 2 children did not take part in the PLS test at 18 months and 3 other children did not take part in PLS assessment at 24 months. Out of the 10 children who were too old for this analysis, 2 parents did not complete the gesture subscale of the UK-CDI at 18 months and 1 child did not take part in the PLS assessment at 18 months and another child did not take part in the assessment at 24 months.

show the different cut-offs points per age group and category (at 12 months for Gestures: 25th percentile: 16 gestures, 50th percentile: 22 gestures and 75th percentile: 28 gestures). The children from the current data set were put into one of the quartile groups according to the prelimary norm cut-off points depending on the amount of gestures used and the age of the child, see Table 23 below.

We would expect around 16 – 17 children per quartile group if the data was equally distributed across groups and similar to the preliminary norm data. From Table 23, we can say that the data was similar to the norm data as our data was not consistently over-represented in one group. This similarity between our sample and the norm sample may help to generalize our findings to the UK population.

Table 23. *Distribution of children into 4 ability groups per UK-CDI measure by percentile ranks (according to preliminary norms)*

	Children (N = 67) distributed into 4 ability groups per measure (using percentile ranks)			
Measures	1st-25th (low ability)	26th-49th (low-average ability)	50th-74th (average-high ability)	75th-99th (high ability)
Gestures 12m	11	15	17	24
Comprehension 12m	17	23	17	10
Production 12m	17	15	20	15
Gestures 18m	12	18	19	18
Comprehension 18m	22	16	15	14
Production 18m	15	21	16	15

There is a certain overlap and movement between the groups. For example at 12 months, out of the 17 children with slow Production development (1st-25th percentile group), seven children also had small Comprehension skills (of which 3 had slow development in terms of Comprehension and Gesture) and one child also had slow Gesture development but typical Comprehension skills. At 18 months, out of the 15 children with a productive delay at 18 months, 10 also had a comprehension delay (of which seven had a combined gesture and comprehension delay). Furthermore, nine of the children with a Production delay were already in the lowest ability group for Production at 12 months. This means about half the children (52.9%) remained slow developers. This proportion was similar for children with sustained Comprehension delay (58.8%) and somewhat less strong for Gesture delay (36.4%). This shows that many children with a delay at 18 months already had low scores six months earlier. Next, I will examine the stability of group ability separately for the different UK-CDI categories (see Table 23). It is also important to note that slow developers often have problems in more than one domain (e.g., Thal et al., 2013).

6.1.2. Materials

The same language tests were used as described above.

6.1.3. Procedure

The procedure was the same as described above.

6.1.4. Statistical analysis

In the following analyses, the z-score transformed language scores were used as above and where necessary the two-step transformed scores were employed- also transformed into z-scores. Missing cases were removed from the analysis.

A first analysis assessed if the groups differed in terms of gender or maternal education (see data in Appendix 12). Chi-square tests of independence were used separately for gender and maternal education with the separate UK-CDI scores (Comprehension, Production and Gestures) at 12 and 18 months. There were no statistically significant differences between the ability groups in terms of gender. Level of maternal education (i.e. low (no qualification to A Level or similar) or high (University degree or similar)) was used as categorised by Fenson et al. (2007). There were no statistically significant differences between the ability groups in terms of maternal education in our data (see Appendix 12 for results).

Multivariate analyses of variance (MANOVAs) were employed to assess if communication ability level had a significant effect on earlier, concurrent or later language scores[42]. MANOVAs were chosen as the data met the assumptions for this type of analysis. The analyses include a range of dependent variables (N = 15, see Table 24 below) and due to the limited sample size it was decided not to compromise the analysis by including covariates.

Table 24. *Dependent variables for different MANOVAs*

Dependent variables
1. Comprehension 12m
2. Production 12m
3. Gestures 12m
4. Comprehension 18m
5. Production 18m
6. Gestures 18m
7. PLS Auditory Comprehension 18m
8. PLS Expressive Communication 18m
9. Comprehension 24m
10. Production 24m
11. Sentence Complexity 24m
12. PLS Auditory Comprehension 24m
13. PLS Expressive Communication 24m
14. ASQ Communication 36m
15. Production 36m
16. Sentence Complexity 36m

In the following, one of the six independent variables of ability group (i.e. ability group of Gestures at 12 and 18 months, ability group of Comprehension at 12 and 18 months and ability group of Production at 12 and 18 months) was used per MANOVA. The dependent variables were the same for all MANOVAs (see Table 24 above), except for the one variable from which the individual ability subgroup (see Table 23 above) was derived (e.g. the continuous variable of

[42] Future analysis could use multilevel/mixed effects models for this repeat-measures design. This would be worthwile as children could be included in the analyses even if they did not take part in all testing stages as these models can account for dropout bias.

Gestures at 12 months was removed when the independent variable of Gesture ability group at 12 months was used).

The MANOVA was then followed-up with Tukey HSD multiple comparisons to test which groups (1^{st}-25^{th} percentile, 26^{th}-49^{th} percentile, 50^{th}-74^{th} percentile, 75^{th}-99^{th} percentile) were significantly different from each other. Some variables were not normal (Expressive Communication at 18 months, Sentence Complexity at 24 and 36 months and ASQ at 36 months) even after 2-step transformation but they were still included in the MANOVA.

6.2. RESULTS

6.2.1. Production

A MANOVA was conducted with Production ability at 12 months (IV) and the different communication variables (DV), see Table 24 except for Production at 12 months. The ability to produce words at 12 months as defined by Production scores (see Table 23 above) had a significant effect on concurrent and later language scores (Pillai's trace $V = 1.0$, $F\,(45,153) = 1.71$, $p = .009$), see Appendix 13. However, separate univariate ANOVAs of ability group by single language measures showed that ability to produce words at 12 months had a significant effect on many dependent variables which demonstrated large effect sizes, as shown in overview in Table 25 below.

Production ability at 12 months had a significant effect on Comprehension scores but not on Gesture knowledge at the same age. Furthermore, Production ability group at 12 months also had a significant effect on Comprehension at 18 and 24 months and Production scores at 18, 24 and 36 months. In addition, these early Production ability groups had an effect on expressive language six months later and even on grammar scores (Sentence Complexity) up to two years later.

Table 25. *Separate univariate ANOVAs between ability groups at 12 months (defined from Production at 12 months) and the communication variables (see Appendix 13)*

Independent variable		F	Sig.	Partial Eta Squared
Ability groups for Production at 12 months	Comprehension 12m	$F(3,63) = 8.29$	.000	.283
	Gestures 12m	$F(3,63) = 1.40$	.250	.063
	Comprehension 18m	$F(3,63) = 8.04$	.000	.277
	Production 18m	$F(3,63) = 7.62$	.000	.266
	Gestures 18m	$F(3,63) = 0.93$	.432	.042
	PLS Auditory Comprehension 18m	$F(3,63) = 1.12$	.347	.051
	PLS Expressive Communication 18m	$F(3,63) = 3.65$	.017	.148
	Comprehension 24m	$F(3,63) = 5.95$	.001	.221
	Production 24m	$F(3,63) = 3.34$	.025	.137
	Sentence Complexity 24m	$F(3,63) = 5.97$	.001	.221
	PLS Auditory Comprehension 24m	$F(3,63) = 0.53$	.661	.025
	PLS Expressive Communication 24m	$F(3,63) = 2.43$	.074	.104
	ASQ communication 36m	$F(3,63) = 0.96$	.418	.044
	Production 36m	$F(3,63) = 3.40$	.023	.139
	Sentence Complexity 36m	$F(3,63) = 6.33$	.001	.232

Tukey post-hoc tests revealed that children in the highest ability group of Production at 12 months had significantly higher Comprehension scores at 12 months (between lowest ability group and highest ability group, $p < .001$; between low-average ability group and highest ability group, $p = .003$; between high-average ability group and highest ability group, $p = .002$)[43] and 18 months than the children in any of the other groups and the other three groups did not differ significantly between each other.

Furthermore, children in the highest Production ability group at 12 months continued to have significantly higher Production scores at 18 months and Comprehension scores at 24 months compared to the two lowest groups but the highest ability group was not significantly different from the average-high group. The lowest ability children did not differ significantly from the children in the two average ability groups.

[43] Please refer to Appendix 13 for the significance levels of the subsequent Tukey HSD tests

Early precocious talkers still had significantly higher scores than children with slow initial development (1^{st}-25^{th} percentile) in terms of Expressive Communication scores six months later (at 18 months), grammar scores one year later (at 24 months) and higher Production scores up to 2 years later (at 24 and 36 months). The average groups did not significantly differ from any of the other groups.

Interestingly, the lowest ability group for Production at 12 months had significantly lower grammar scores at 36 months than all other children. The other three groups did not differ in terms of grammar scores.

We expected that language was more stable at 18 months compared to 12 months, therefore we calculated the same MANOVA with Production ability at 18 months (IV) and the communication variables (IV, see Table 24 except for language variables at 12 months and Production at 18 months) in the next step. Again, a MANOVA was used and the result was significant (Pillai's trace V = 1.02, F (36,162) = 2.33, p < .001), see Table 26 for the separate ANOVAs.

Table 26. *Separate univariate ANOVAs between Production ability at 18 months and the communication variables*

Source		F	Sig.	Partial Eta Squared
Levels of ability for Production 18m	Comprehension 18m	$F(3,63) = 16.88$	.000	.446
	Gestures 18m	$F(3,63) = 5.38$	.002	.204
	PLS Auditory Comprehension 18m	$F(3,63) = 7.32$	.000	.258
	PLS Expressive Communication 18m	$F(3,63) = 25.74$	.000	.551
	Comprehension 24m	$F(3,63) = 18.99$	.000	.475
	Production 24m	$F(3,63) = 19.49$	.000	.481
	Sentence Complexity 24m	$F(3,63) = 17.09$	.000	.449
	PLS Auditory Comprehension 24m	$F(3,63) = 2.42$	.075	.103
	PLS Expressive Communication 24m	$F(3,63) = 10.33$	.000	.330
	ASQ communication 36m	$F(3,63) = 5.06$	.003	.194
	Production 36m	$F(3,63) = 7.34$	.000	.259
	Sentence Complexity 36m	$F(3,63) = 4.86$	.004	.188

Separate ANOVAs showed that Production ability at 18 months had a significant effect on most communication variables (except for Gestures at 12 months and Auditory Comprehension at 24 months). The effect sizes were large.

The significant relationships were further analysed using post-hoc Tukey tests. The lowest and highest groups were significantly different from each other in all situations; this means children who were classified as late talkers (production at or below 25th percentile) at 18 months already showed significantly lower concurrent communication scores and showed significant lower vocabulary, general language and grammar scores up to 18 months later (at 36 months) when compared to those high-achieving children at 18 months.

There was no scenario were all groups differed significantly from each other. In addition, groups were less likely to differ if they were of neighbouring ability at 18 months (e.g. the two middle groups often did not differ significantly from each other). For Production at 18 months the high-level group was significantly different (higher) from the other three groups in terms of Comprehension and Production at 24 months. In contrast, the lowest scoring group was not significantly different from the low-medium group in those two situations.

At 36 months, only the high and low ability group differed significantly from each other in terms of Sentence Complexity. At this point, Production was still the best outcome variable to distinguish the children from each other; however, the low ability group was not significantly different from the low-medium ability group and the high ability group was not different from the high-medium ability group (again the two middle groups did not differ significantly from each other). As expected with regards to the ASQ, the high ability group was not significantly different compared to the middle ability groups (due to the purposefully created negative skew when constructing the tool); however, the low ability group was still of significantly lower ability regarding the middle-high and high ability group but not the low-medium ability group.

6.2.2. Comprehension

A MANOVA was used with Comprehension ability at 12 months (IV) and the different communication variables (DV), see Table 24 except for Comprehension at 12 months. The analysis showed that Comprehension ability at 12 months had a significant effect on communication scores between 12 and 36 months (Pillai's trace *V* = 1.08, *F* (45,153) = 1.91, *p* = .002), see Table 27 for the results of the separate ANOVAs.

Table 27. *Separate univariate ANOVAs between comprehension ability at 12 months and the communication variables*

Source		F	Sig.	Partial Eta Squared
Levels of ability for Comprehension at 12m	Production 12m	*F*(3,63) = 9.08	.000	.302
	Gestures 12m	*F*(3,63) = 2.00	.125	.086
	Comprehension 18m	*F*(3,63) = 12.52	.000	.373
	Production 18m	*F*(3,63) = 3.85	.014	.155
	Gestures 18m	*F*(3,63) = 1.22	.309	.055
	PLS Auditory Comprehension 18m	*F*(3,63) = 3.65	.017	.148
	PLS Expressive Communication 18m	*F*(3,63) = 2.85	.045	.119
	Comprehension 24m	*F*(3,63) = 11.91	.000	.362
	Production 24m	*F*(3,63) = 5.41	.002	.205
	Sentence Complexity 24m	*F*(3,63) = 1.36	.264	.061
	PLS Auditory Comprehension 24m	*F*(3,63) = 1.76	.164	.077
	PLS Expressive Communication 24m	*F*(3,63) = 2.72	.052	.115
	ASQ Communication 36m	*F*(3,63) = 2.94	.040	.123
	Production 36m	*F*(3,63) = 2.91	.041	.122
	Sentence Complexity 36m	*F*(3,63) = 1.50	.223	.067

The data indicates that very early Comprehension (12 months) ability had a significant effect on several communication variables up to 36 months showing mostly large effect sizes. There were also some ANOVAs with non-significant results (i.e. Gestures at 12 and 18 months, Sentence Complexity at 24 and 36 months, PLS scores at 24 months).

Post-hoc tests showed that the Comprehension ability group at 12 months had a significant effect on Production scores at 12 months. Children from the highest ability group had significantly higher scores compared to all other three groups which showed lower Production skills and they did not differ from each other. In addition, Comprehension ability at 12 months had a significant effect on Comprehension scores at 18 and 24 months. Whilst the neighbouring scores at the extreme ends (between the low and low-average ability groups well as between the average-high and high ability groups) were not significantly different from each other, all the other groups did significantly differ from each other.

In addition, Comprehension ability at 12 months also had a significant effect on Production scores (18, 24 and 36 months). At 18 months, children from the highest ability group had significantly higher scores compared to the two lowest groups only. The other groups were not significantly different from each other. At 24 months, the low-average ability group showed significantly different scores from the highest and the average-high group. The other groups did not differ significantly which was interesting particularly for the lowest ability group which consistently showed significantly lower scores in terms of Production ability (see MANOVAs above). Similar results were found at 36 months, only the low-average ability group was significantly different from the highest ability group in terms of Production scores and only the two average ability groups differed in terms of ASQ scores, again the other groups were not statistically different from each other.

Another MANOVA was used with Comprehension ability at 18 months (IV) and the communication variables (IV, see Table 24 except for language variables at 12 months and Comprehension at 18 months). The model as a whole had a significant effect on communication scores between 12 and 36 months (Pillai's trace V = 1.16, $F(36,162) = 2.85$, $p < .001$). Furthermore, Comprehension ability at 18 months had a significant effect on all scores except for the ASQ at 36 months (see Table 28 for the results of the separate ANOVAs), this contrasts with Comprehension ability at 12 months (see Appendix 13 for table of results). The effect sizes were large for all significant ANOVAs.

Table 28. *Separate univariate ANOVAs between Comprehension ability at 18 months and the communication variables*

Source		F	Sig.	Partial Eta Squared
Levels of ability for Comprehension 18m	Production 18m	$F(3,63) = 12.43$	.000	.372
	Gestures 18m	$F(3,63) = 13.73$	.000	.395
	PLS Auditory Comprehension 18m	$F(3,63) = 11.95$	.000	.363
	PLS Expressive Communication 18m	$F(3,63) = 8.08$	.000	.278
	Comprehension 24m	$F(3,63) = 26.92$	.000	.562
	Production 24m	$F(3,63) = 14.72$	.000	.412
	Sentence Complexity 24m	$F(3,63) = 7.84$	.000	.272
	PLS Auditory Comprehension 24m	$F(3,63) = 5.38$	.002	.204
	PLS Expressive Communication 24m	$F(3,63) = 5.04$	.003	.193
	ASQ Communication 36m	$F(3,63) = 2.45$	.072	.104
	Production 36m	$F(3,63) = 12.77$	.000	.378
	Sentence Complexity 36m	$F(3,63) = 3.61$	.018	.147

Post hoc tests revealed that Comprehension at 18 months had a significant effect on all Gestures, vocabulary (Comprehension and Production), general language scores (PLS-5 UK) and grammar scores (except for the ASQ) in which high ability and low ability groups were significantly different from each other. Communication scores were always significantly lower for the low ability group compared to the high ability group.

Comprehension level (18 months) was best at distinguishing between Production scores (18 months) with the high Comprehension ability group showing significantly higher Production scores than the other three groups. Within these groups, only the lowest and average-high group also differed significantly demonstrating significantly lower Production scores in the lowest Comprehension ability group compared to the average-high group. At 18 months, children in the lowest Comprehension group also had significantly fewer Gestures than all the other groups (the other three groups did not differ significantly). In terms of general language scores on the PLS, significant differences were only found for more distant but rather than neighbouring Comprehension ability groups.

Furthermore, Comprehension ability showed significantly different scores in terms of Comprehension scores (at 24 months), whilst the average-high and high ability groups were significantly different from all other groups (and remained in the same order), the two lowest groups were not significantly different from each other. Furthermore, children in the highest Comprehension ability group at 18 months showed superior Production scores at 24 and 36 months, the three lower groups did not differ from each other. Only distant groups showed significant differences in terms of general language scores at 24 months (low and high ability groups, low-average and high ability groups, low and average-high ability groups) and grammar scores at 24 months (low and high ability groups, low-average and high ability groups) and 36 months (low and high ability groups).

6.2.3. Gestures

When running the MANOVA with Gesture level at 12 months or 18 months (IVs) and the communication variables (DVs), the models were not significant (see Appendix 13 for overview of results).

6.3. DISCUSSION

In this analysis, children were allocated into one of four ability groups for Production, Comprehension and Gestures at 12 and 18 months. Four ability groups were created from the preliminary UK-CDI norm data (Alcock et al., 2017) in terms of percentile scores for low (1^{st}-25^{th} percentiles), low-average (26^{th}-49^{th} percentiles), average-high (50^{th}-74^{th} percentiles) and high (75^{th}-99^{th} percentiles) skills.

Children in the high ability groups for Comprehension and Production at 12 and 18 months continued to have significantly higher scores than low ability children in terms of a variety of different linguistic categories up to 36 months of age. Furthermore, particularly for Production ability at 18 months, average scores indicated that the four ability groups remained in the same position across the different communication variables. Nevertheless, the two average groups were not significantly different in terms of any of the communication variables. In contrast, the Comprehension ability groups at 12 and 18 months showed that the low ability children had higher scores than the low-average children on average for several variables, but these differences were not significant. This indicates more variability in outcome among the low ability children compared to the high ability children.

We expected that high ability children continued to have higher scores compared to average and low ability children. Our data confirms the hypothesis but shows at the same time that neighbouring groups (e.g. high ability and average-high ability groups) were less likely to differ from each other compared to more distant ability groups (e.g. high ability and low-average ability groups).

Furthermore, in many cases children in the low ability group did not significantly differ from the low-average group which means there was more change over time in the low and low-average groups (particularly for Comprehension ability groups). Even though it should be noted that in several cases the high ability group was also not significantly different from the average-high ability children, the high ability group remained more stable compared to the lower ability groups. Similarly, Thal et al. (1997) also indicated that early-talkers (at or above the 90^{th} percentile) compared to late-talkers (at or below the 10^{th} percentile) remained more stable in their ability between 13 and 26 months. This is because most children with slow initial Comprehension or Production development catch up with their peers by kindergarten age when tested on a range of standardised language assessments (Thal et al, 2013).

However, in terms of grammar the lowest ability group for Production at 12 months had significantly lower grammar scores (Sentence Complexity) at 36 months compared to all other groups in the current study. The high, average-high and low-average ability groups were not significantly different from each other. This result may have to be interpreted with caution as other authors suggested that early vocabulary below 13 months is not sufficiently developed in order to make predictions for later language (Thal et al., 1997; Thal et al., 2013).

This leads to the next hypothesis which was investigated, namely whether there was more continuity for the same language categories compared to different categories. We found that especially for Production ability at 18 months there was good stability with Production scores at 24 and 36 months. However, there was even stronger continuity between Production ability at 18 months and grammar at 24 months (Sentence Complexity). Children with low Production ability had significantly lower grammar scores compared to all other groups and the high ability children had significantly higher grammar scores compared to all other groups. Again, these results show that children with low Production ability at 18 months show more variability in outcome in terms of Production as they did not significantly differ from the low-average group at any age point at 24 and 36 months. This indicates that they may catch up with their peers over time (as suggested by Thal et al., 2013).

However, children with low Production at 18 months differed significantly from all other groups at 24 months in terms of grammar and from high and average-high ability children in terms of the ASQ communication subscale (primarily assessing the understanding and use of grammar) at 36 months. This shows that children with slow Production at 18 months catch up with their peers in terms of vocabulary first but at the same time show early grammar delays at 24 months. At 36 months, the data suggests that many of these children also start to catch up in terms of grammar. This fits well with previously reported studies that those late-talkers who catch up with their peers show an increase in vocabulary development but often demonstrate protracted grammar delays (e.g., Rescorla, Roberts, & Dahlsgaard, 1997; Rescorla, 2011).

This developmental trajectory describes developmental delay (i.e. late onset and/or slow rate of development) rather than atypical development (i.e. deviant development compared to chronologically age matched and mentally age matched children) (Thomas et al., 2009). Furthermore, these results support the notion that language delay and language disorders are not distinct phenomena, but are linked as problems in later acquired areas (e.g. literacy skills) are associated with earlier deficits in language components acquired during previous stages of development (see also Thal et al., 2013).

In addition, we found that the prediction of scores was better for Comprehension and Production than for Gestures. Our data shows that children with low Comprehension or Production ability at 18 months also had significantly lower Gesture scores at 18 months compared to the other three groups which did not differ significantly from each other. Similar to our results, Thal et al. (2013) showed that late comprehenders used significantly fewer Gestures between 10 and 16 months compared to late producers and typically developing children. Late producers also used significantly fewer Gestures than typically developing children. Other research has shown that Gestures at 13 months were able to predict late-talking status (Production scores below the 10th percentile) but not early talking status (Production scores above the 90th percentile) at 20 months (Thal et al., 1997). Similarly, we found that children with low Comprehension and Production ability at 18 months used significantly fewer types of Gestures than the other ability groups. At the same time, Gesture scores did not differ between average and early talkers (high Production ability) at 18 months. This shows that Gestures are important for early language development but the influence on later language becomes weaker as shown by Hsu and Iyer (2016) and Rose et al. (2009). Gestures at 15 months had an indirect effect on language impairment (LI) risk at 3 and 4;6 years because the influence of early Gestures on later LI risk was mediated by Production at 15 months (Hsu & Iyer, 2016). Furthermore, Rose et al. (2009) found that in contrast to Comprehension and Production, Gestures at 12 months did not add any significant variance to vocabulary comprehension or verbal fluency at 36 months. This shows that whilst Gestures helped to predict late-talking status up to 20 months, they could not predict language beyond three years of age. As found in this research, Hsu and Iyer (2016) and Rose et al. (2009) could show that Comprehension and Production rather than Gestures need to be used to significantly predict language at 3 years and beyond. Furthermore, we found that language stability was better with ability groups (for Comprehension and Production) at 18 months compared to 12 months. This is also supported by previous research which found stronger relationships when the children were older at the first time of testing (Cochet & Byrne, 2016; Friend et al., 2012). Less strong correlations between language variables at 12 months with later language may be explained by the neural reorganization which takes place around the child's first birthday (McCall et al., 1977).

In sum, the current study suggests that there is stability in early language. Especially high ability children remained at an elevated level compared to the other groups, and especially compared to the low ability children. Overall, we found that late-talkers (low Production ability at 18 months) remained below the children with average or above average ability (average-high and high ability groups) in regards to all language areas when followed-up at 24 and 36 months (except for Sentence Complexity at 36 months[44]). Production ability at 18 months showed similar but somewhat better stability with later language than Comprehension ability at 18 months. In addition, both ability groups at 18 months showed better stability than at 12 months. Particularly for the low Production ability groups stability was strongest with grammar and not Production or any other language variable.

[44] Here, the low ability group only differed significantly from the high ability group but not from the average-high and low-average groups.

7. ANALYSIS 3: FAMILY, BIOLOGICAL, ENVIRONMENTAL AND EARLY COMMUNICATION INFLUENCES ON LANGUAGE DEVELOPMENT

Children's language development is influenced by different family, biological and environmental factors as pointed out in the introduction. It has already been established above (see Chapter 5 and 6) that there were no CDI differences regarding preterm birth and few gender differences favouring girls in terms of CDI and PLS scores in the present sample.

The following analyses investigate further factors which may influence language development, namely sibling status, family risk of dyslexia and speech and language problems, cognitive skills. The aim of the following analysis was to establish how much variance early language scores (UK-CDI scores) could explain in later language scores. It was then evaluated if UK-CDI scores at 12 months could make significant contributions to later language scores above and beyond UK-CDI scores at 18 months. This may be important for the use in practice – if UK-CDI scores at 12 months do not explain any additional significant variance in later language and if UK-CDI scores at 18 months explain sufficient language skills at later stages it may be sufficient to only collect UK-CDI information at 18 months without collecting any UK-CDI information at 12 months.

It will also be examined how much additional variance environmental, family and biological factors contribute towards later language scores above and beyond early communication scores (UK-CDI at 12 and 18 months). This is important as research suggests that the environment may have a greater influence on early language development and genetic factors within families may explain more variance in later language ability (Hayiou-Thomas et al., 2012). If the UK-CDI is used for assessing UK children's language status in the future, it could be important to know in which way UK-CDI scores can predict later ability and if other factors (e.g. gender, sibling status) significantly contribute to these later scores beyond the UK-CDI scores.

7.1. METHOD

7.1.1. Participants

Participant numbers were only slightly different depending on the tests used as some children did not take part in all in-person assessments (see Table 29); however, data was available for all children for the CDI-type questionnaires between 12 and 36 months and the Family Questionnaire (N= 82). For information about gender and age see Chapter 5 (5.1.1. Participants).

Table 29. *Number of participants used per instrument*

Type of instrument	N
PLS 18m	79
PLS 24m	78
Bayley cognitive 18m	79
Bayley cognitive 24m	79
Bayley motor 18m	78
Bayley motor 24m	78
ASQ 36m	82

7.1.2. Materials

The same instruments were used as in the previous studies. In addition, the Bayley Scales of Infant and Toddler Development, Third Edition (Bayley-III). The Bayley has norms for UK children and can be used between 1 and 42 months to assess cognitive, language and (fine and gross) motor skills in-person by a trained examiner. For this research, I used the cognitive and fine and gross motor scales. Data collection took place at 18 and 24 months of age. This is important as previous research has shown that language acquisition is connected to motor (e.g., Iverson, 2010; Leonard & Hill, 2014; Longobardi, Spataro, & Rossi-Arnaud, 2014) and cognitive (e.g., Miller & Marcovitch, 2015; Rose et al., 2009) development.

Psychometric properties of the Bayley-III were assessed by the original research team (Bayley, 2006b). The average reliability coefficients measuring internal consistency varied between .86 - .93 across the different in-person subscales. Test-retest reliability was assessed for all in-person subscales. The correlations varied between .67 - .94, showing higher correlations with increasing age. For the adaptive behaviour scale, test-retest coefficients were between .71 and .92. Inter-rater reliability for the adaptive behaviour scale showed correlations between .59 - .86. Concurrent validity was high (r = .72 - .79) for the cognitive subscale in comparison to the IQ measures of the WPPSI-III (Wechsler, 2002). The motor scale correlated with the PDMS-2 (Folio & Fewell, 2000) for fine motor, gross motor and visual-motor integration scores (r = .55 - .59). Raw scores are standardised to composite scores with a mean of 100 and a standard deviation of 15. Whilst there are different classifications for developmental delay, a frequently used criterion is based on standard deviations below the mean of the norming group (i.e. 2 standard deviations in one area or 1.5 deviations in two or more areas) (Bayley, 2006b). Composite scores of 69 and below identify a child as having a developmental delay whilst children scoring between 70 and 84 as having a mild delay (B. Hutchon, personal communication, February 24, 2015).

7.1.3. Procedure

The procedure was the same as above for the tools explained above. The Bayley Scales were administered at 18 and 24 months. The assessment was conducted as specified in the Bayley manual and the experimenter was professionally trained in a two-day workshop for the Bayley Scales of Infant and Toddler Development (Bayley III). At the beginning of testing parents were made aware that some questions would be easy, and some would be too difficult for the child. They were also told that the test needed to be administered in a certain way and that they should not help their child (i.e. speaking, pointing, eye-pointing, manipulating child's position by leaning in/out or pushing a child's arm into a certain direction). However, it was also made clear that the experimenter would sometimes need the parent's help but that their help would be specifically requested. The experimenter was seated on a chair in front of a table with the parent across the table and their child on their lap. Testing began with the cognitive scale followed by the motor scale. All items were administered at the table. An exception was the gross motor scale which required the child to perform actions whilst standing or sitting on the floor. The aim was to complete all

Bayley scales in one session with some breaks but if children showed signs of distress or tiredness testing stopped. Another date for finishing the testing was then agreed. Reimbursement was the same as mentioned above.

7.1.4. Statistical analysis

Hierarchical multiple regressions were used in the following. It was first investigated how much variance UK-CDI scores (i.e. Comprehension, Production, Gestures and Phrases Understood) could contribute to later language scores. This was calculated for UK-CDI scores at 12 months and at 18 months separately. As we only had a limited sample size, we limited the number of predictor variables to a maximum of five per regression analysis to not exceed a limit of 15 participants per predictor (Stevens, 1996). The sample size varied across analyses depending on the background variables (e.g. all parents provided information about the child's gender but not all parents gave information about their household income), see Table 16 above.

The assumptions for hierarchical multiple regressions were checked prior to conducting the analyses. The dependent variables were the same across all regressions, see Table 30 below.

Table 30. *Dependent variables (outcome variables) used in individual hierarchical multiple regressions*

1. PLS total 18m
2. Production 18m
3. PLS total 24m
4. Production 24m
5. ASQ 36m
6. Production 36m
7. Sentence Complexity 36m

Two variables were not normally distributed (ASQ and Sentence Complexity at 36 months). Pallant (2013) advises to use more participants in such cases, thus the number of five predictors was not exceeded in any of the following calculations.

For the first regressions, the UK-CDI predictor variables (independent variables) were Comprehension, Production, Gestures and Phrases Understood at 12 months and the control variable was age at 12 months. The regression was run seven times with the different dependent variables.

For the second regression, the same predictor variables were used at 18 months and age at 18 months was used as the control variable. Five regressions were run for the dependent variables collected at 24 and 36 months.

Next, the most predictive UK-CDI scores (as defined from the previous regression analyses) were selected and used as the control variables. The predictor variables were the family, biological and environmental factors introduced in this chapter (i.e. prematurity, gender, SES, sibling status, ear infections, sleep, family risk status of dyslexia or speech or language problems, cognitive and motor skills). These predictor variables were used individually with the same dependent variables as in Table 30 to find out how much variance they explained above and beyond the UK-CDI scores. In a next analysis, they were used in the first step followed by CDI variables in the next step to find out how much variance they could explain on their own before CDI scores were added.

Further relationships between the family and potentially hereditary factors and environmental influence on language development were examined. Depending on the type of data (continuous versus nominal data), possible associations were tested using Spearman’s rho tests and to control for inflation of type 1 errors, Bonferroni correction was used.

Furthermore, group differences were investigated using MANOVAs or MANCOVAs to test the effect of family, hereditary and environmental factors on language development. The independent variables were the specific family, biological or environmental factor (e.g. maternal education) and the dependent variables were the 16 communication variables tested between 12 and 36 months (see Table 31) and for MANCOVAs, age was the covariate. In analyses with missing data, participants were removed; however, their data was used in other analyses if the required data was available. Furthermore, in analyses with very different group sizes, the means were weighed to balance unequal groups.

If group differences or associations were present at two or more different points between 12 and 36 months, the data was further analysed. For example, it was investigated if family, biological or environmental factors had a moderating (interaction) effect on the relationships between early UK-CDI scores and later language ability. This is important as it is possible that the association between early and later language related skills is dependent on a third (moderator) variable. An example is the study by Spinelli et al. (in press) which hypothesised that mother's quality of language input between 6 and 12 months was a significant moderator for the longterm association between children's attentional abilities at 3 months and language outcomes at the end of the 2^{nd} year. Indeed, those children with greater attentional skills and more complex and varied maternal input had better language skills at 18 and 24 months of age. In my study, it is possible that variables such as hours of sleep or motor skills could be moderators between early and later language skills and hence will be investigated where applicable.

Table 31. *Dependent variables for different MANOVAs/MANCOVAs and Spearman's rho correlations*

Dependent variables
Phrases Understood (12 months)
Comprehension (12 months)
Production (12 months)
Gestures (12 months)
Phrases Understood (18 months)
Comprehension (18 months)
Production (18 months)
Gestures (18 months)
PLS Total Language (18 months)
Comprehension (24 months)
Production (24 months)
Sentence Complexity (24 months)
PLS Total Language (24 months)
ASQ Communication (36 months)
Production (36 months)
Sentence Complexity (36 months)

7.2. RESULTS

7.2.1. Does the UK-CDI at 12 months predict later language scores?

A hierarchical multiple regression was used to assess if UK-CDI scores at 12 months (IV- Comprehension, Production, Gestures and Phrases Understood) predict PLS scores at 18 months (DV), after controlling for age at 12 months (covariate). Age at 12 months was entered as step 1, explaining 3% of the variance in PLS scores at 18 months. After the entry of the four CDI variables at 12 months, the total variance explained by the model as a whole was 14.6%, F (5, 76) = 2.60, p = .032. The UK-CDI scores explained an additional 11.6% of the variance in general language ability six months later (PLS 18 months), after controlling for age when the UK-CDI was completed, R squared change = .116, F change (4, 76) = .2.59, p = .043. In the final model, only age at 12 months was significant (*beta* = -.33, p = .008).

Preliminary analysis found that Comprehension and Phrases Understood correlated strongly (r = .79), hence violating the assumption of multicollinearity. However, after removing Phrases

Understood from the analysis, the results remained very similar. The total variance explained by the model was again 14.6%, F (4, 77) = 3.28, p = .015. The only significant predictor was still age at 12 months (*beta* = -.33, p = .007), in the final model.

The same hierarchical multiple regressions were conducted with Production at 18 months, PLS at 24 months, Production at 24 months, ASQ at 36 months, Production at 36 months and Sentence Complexity at 36 months, see Table 32. Again, the result did not change after removing Phrases Understood from the analysis.

Table 32. *Results of the hierarchical multiple regression analysis predicting from UK-CDI scores at 12 months*

Outcome	step predictor	*t* entry	*t* final	*B*	*SE B*	*β*	R^2 step	Δ R^2	*F* change
Production 18m									
	1. Age at 12m	.45††	-2.08*	-.18	.09	-.21	.002		
	2. UK-CDI scores at 12m						.40	.395	12.47***
	Comprehension		-.70††	-.11	.16	-.11			
	Production		4.78***	.56	.12	.54			
	Gestures		2.30*	.30	.13	.29			
	Phrases Understood		.40††	.06	.15	.06			
PLS 24m									
	1. Age at 12m	-1.02††	-1.96†	-.20	.10	-.24	.01		
	2. UK-CDI scores at 12m						.09	.08	1.69††
	Comprehension		.91††	.17	.18	.18			
	Production		.43††	.06	.14	.06			
	Gestures		.95††	.14	.15	.14			
	Phrases Understood		.12††	-.02	.18	-.02			
Production 24m									
	1. Age at 12m	-.30††	-1.67†	-.18	.10	-.20	.001		
	2. UK-CDI scores at 12m						.149	.148	3.30*
	Comprehension		.83††	.16	.19	.16			
	Production		1.68†	.24	.14	.23			
	Gestures		.97††	.15	.15	.24			
	Phrases Understood		-.17††	-.03	.18	-.03			

ASQ 36m										
	1.	Age at 12m	-.55††	-1.12††	-.10	.09	-.14	.004		
	2.	UK-CDI scores at 12m						.09	.086	1.80††
		Comprehension		-.40††	-.07	.16	-.08			
		Production		.94††	.11	.12	.13			
		Gestures		.05††	.01	.13	.01			
		Phrases Understood		1.44††	.23	.16	.27			
Production 36m										
	1.	Age at 12m	-1.22††	-1.99*	-.20	.10	-.25	.02		
	2.	UK-CDI scores at 12m						.10	.084	1.78††
		Comprehension		.53††	.10	.18	.10			
		Production		.86††	.11	.13	.12			
		Gestures		.42††	.06	.15	.06			
		Phrases Understood		.46††	.08	.17	.08			
Sentence Complexity 36m										
	1.	Age at 12m	-.37††	-1.24††	-.11	.09	-.15	.002		
	2.	UK-CDI scores at 12m						.13	.12	.269*
		Comprehension		-.10††	-.02	.17	-.02			
		Production		2.10*	.26	.12	.29			
		Gestures		.11††	.01	.13	.02			
		Phrases Understood		.70††	.11	.16	.13			

[a] Standardised beta at each step.

* p < 0.05; ** p < 0.01; *** p <0.001; p > 0.05†; p > 0.1††

To sum up all results, we found that whilst CDI scores at 12 months were of limited use to predict future language scores, Production and Gestures were the best predictors. This was somewhat surprising as the average age of the children was 11.71 months (*SD* 1.19) at 12 months which would indicate that most children were only starting to say their first words, hence it was expected that comprehension and gestures were better predictors at this age.

7.2.2. Does the UK-CDI at 18 months predict later language scores?

A hierarchical multiple regression was used to assess if UK-CDI scores at 18 months (IV-Comprehension, Production, Gestures and Phrases Understood) predict PLS scores at 24 months (DV), after controlling for age at 18 months (covariate). Further hierarchical multiple regressions were used with other language variables at 24 and 36 months (IVs) and the same independent variables and covariate (see Table 33). There were no multicollinearity issues at 18 months between Comprehension 18 months and Phrases Understood 18 months, hence all UK-CDI variables at 18 months were included as predictors (Step 1: Age at 18 months, Step 2: Comprehension at 18 months, Production at 18 months, Gestures at 18 months, Phrases Understood at 18 months) and the same outcome variables were used as in the example above except for Production and PLS at 18 months. As can be seen from Table 33 below, UK-CDI variables at 18 months were better than at 12 months for predicting future language scores. At this age, Comprehension and Production scores were better than scores for Gestures or Phrases Understood to make predictions about language development up to 24 and even 36 months of age.

Table 33. *Results of the hierarchical multiple regression analysis predicting from UK-CDI scores at 18 months*

Outcome	step predictor	*t* entry	*t* final	*B*	*SE B*	*β*	R^2 step	Δ R^2	*F* change
PLS 24m									
	1. Age at 18m	-1.15††	-1.12††	-.17	.16	-.12	.02		
	2. UK-CDI scores at 18m						.34	.32	9.35***
	Comprehension		1.95†	.32	.17	.33			
	Production		1.88†	.24	.13	.25			
	Gestures		.40††	.05	.11	.05			
	Phrases Understood		.10††	.01	.14	.01			
Production 24m									
	1. Age at 18m	-.84††	-1.11††	-.16	.14	-.10	.01		
	2. UK-CDI scores at 18m						.51	.50	19.18***
	Comprehension		2.58*	.40	.15	.38			
	Production		3.93***	.46	.12	.46			
	Gestures		-.51††	-.05	.10	-.05			
	Phrases Understood		-.50††	-.06	.13	-.06			
ASQ 36m									
	1. Age at 18m	-1.02††	-1.55††	-.23	.15	-.17	.01		
	2. UK-CDI scores at 18m						.27	.26	6.66***
	Comprehension		.09††	.01	.16	.02			
	Production		2.86**	.34	.12	.41			
	Gestures		-.29††	-.03	.11	-.04			
	Phrases Understood		1.28††	.17	.13	.18			

Production 36m									
	1. Age at 18m	-1.23††	-.90††	-.13	.15	-.09	.02		
	2. UK-CDI scores at 18m						.41	.39	12.45***
	Comprehension		3.06**	.48	.16	.50			
	Production		1.98†	.24	.12	.25			
	Gestures		-1.08††	-.12	.11	-.12			
	Phrases Understood		-.12	-.02	.13	-.02			
Sentence Complexity 36m									
	1. Age at 18m	.24††	-.04††	-.01	.15	-.01	.01		
	2. UK-CDI scores at 18m						.29	.28	7.55***
	Comprehension		.48††	.08	.16	.09			
	Production		2.83**	.34	.12	.40			
	Gestures		-1.73†	-.19	.11	-.21			
	Phrases Understood		1.71††	.23	.13	.24			

[a] Standardised beta at each step.

* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$; $p > 0.05$†; $p > 0.1$††

It was expected that Production at 18 months predicted language scores up to 36 months, as there was a large variability in Production scores at 18 months compared to 12 months and the associations were also strong between Production at 18 months and later language, see correlations in Analysis 1. Our data also showed that Comprehension at 18 months predicted later Production at 36 months. This result confirms previous reports about the relationship between vocabulary comprehension and later vocabulary production (e.g., Thal et al., 2013).

Does the UK-CDI at 12 months predict language scores beyond the UK-CDI at 18 months?

The performance of the predictors varied depending on the age of the UK-CDI use. At around 12 months, the best predictors were Production and Gestures in contrast to Comprehension and Production six months later at 18 months. It was further investigated if UK-CDI scores at 12 months contributed significant variance beyond the UK-CDI data at 18 months. The best predictors from the previous analyses were used (Step 1: Comprehension 18 months, Production 18 months; Step 2: Age at 12 months; Step 3: Production 12 months, Gestures 12 months) and again the same outcome variables were employed. Hence, a hierarchical multiple regression used different independent variables (UK-CDI scores at 18 months, age at 12 months, Production at 12 months, Gestures at 12 months) to predict PLS scores at 24 months (DV), please see Table 34 for more analyses with other dependent variables. The results showed that UK-CDI completed at 12 months could not explain any additional variance after controlling for 18 months UK-CDI scores, see Table 34 below. As was expected that Comprehension would generally be a better predictor than Production at 12 months, the same analyses were rerun with Comprehension 12 months and Gestures 12 months as Step 3. However, the same non-significant results were found with these variables.

Table 34. *Results of the hierarchical multiple regression analysis predicting from UK-CDI scores at 18 months*

Outcome	step predictor		*t* entry	*t* final	*B*	*SE B*	*β*	*R²* step	Δ *R²*	*F* change
PLS 24 m										
	1.	UK-CDI scores at 18m						.33		
		Comprehension	3.36***	.34***	.44	.13	.45			
		Production	.1.66††	.2.34*	.30	.13	.32			
	2.	Age at 12m	-1.24††	-.32††	-.03	.09	-.03	.34	.01	1.53††
	3.	UK-CDI scores at 12m						.37	.03	1.54††
		Production		-1.56††	-.18	.12	-.19			
		Gestures		-.60††	-.07	.12	-.07			
Production 24m										
	1.	UK-CDI scores at 18m						.49		
		Comprehension	3.37***	3.53***	.43	.12	.41			
		Production	3.75***	3.90***	.47	.12	.47			
	2.	Age at 12m	-.56††	.29††	.02	.08	.03	.49	.00	.31††
	3.	UK-CDI scores at 12m						.51	.01	1.07††
		Production		-.92††	-.10	.11	-.10			
		Gestures		-.01††	-.11	.11	-.11			
ASQ 36m										
	1.	UK-CDI scores at 18m						.23		
		Comprehension	1.56††	1.76†	.22	.13	.25			
		Production	2.42*	2.56*	.32	.13	.38			

2. Age at 12m	-.74††	.01††	.00	.09	.00	.24	.01	.54††
3. UK-CDI scores at 12m						.25	.01	.66††
Production		-.57††	-.06	.11	-.07			
Gestures		-.91††	-.11	.12	-.12			
Production 36m								
1. UK-CDI scores at 18m						.39		
Comprehension	4.01***	4.35***	.52	.12	.54			
Production	1.66††	2.32*	.28	.12	.30			
2. Age at 12m	-1.53††	-.20††	-.02	.08	-.02	.41	.02	2.33††
3. UK-CDI scores at 12m						.44	.03	2.15††
Production		-1.18††	-.13	.11	-.13			
Gestures		-1.53††	-.17	.11	-.18			
Sentence Complexity 36m								
1. UK-CDI scores at 18m						.24		
Comprehension	1.25††	1.35††	.17	.13	.19			
Production	2.76**	2.27*	.29	.13	.34			
2. Age at 12m	-.56††	-.24††	-.02	.09	-.03	.24	.00	.31††
3. UK-CDI scores at 12m						.25	.01	.55††
Production		.72††	.08	.12	.09			
Gestures		-.85††	-.10	.12	-.11			

[a] Standardised beta at each step.

* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$; $p > 0.05$†; $p > 0.1$††

7.2.3. Contributions of family, biological and environmental factors on the prediction of language

As we know that UK-CDI scores at 12 and 18 months make significant contributions to later language on their own when analysed separately, it was further examined which family, biological and environmental factors helped to predict language after controlling for UK-CDI scores at 12 and 18 months, using the same outcome variables as above. Age was also controlled at 12 but not 18 months, due to the high variation in age at 12 months. For UK-CDI scores at 12 months (Production and Gestures) as the predictor variables, the same analyses were rerun with Comprehension at 12 months and Gestures at 12 months but the results did not differ, thus only the results for Production and Gestures at 12 months were reported in the following.

7.2.3.1. Contributions of prematurity to later language scores

Hierarchical multiple regressions were used with different outcome variables (DV), see Table 30. Depending on the analysis the first step was age at 12 months and the second step was Production and Gestures at 12 months. Alternatively, Production and Comprehension at 18 months were entered as a first step. Prematurity was entered as a last step. After controlling for UK-CDI scores at 12 or 18 months, we found that prematurity did not add any significant contributions to later language scores in our sample. This status also did not contribute any significant variance to later language scores on its own before adding CDI scores as predictors into the model. This result matches the finding in Analysis 1 in which CDI scores did not differ between preterm and full-term children.

7.2.3.2. Contributions of gender to later language scores

The same hierarchical regressions were used as above in 7.2.3.1., but as a last step gender rather than prematurity was entered. Gender contributed significant variance to help predict PLS scores at 24 months beyond UK-CDI scores. Gender but not UK-CDI scores at 12 months contributed significantly (5.8% of variance) to PLS scores at 24 months (*R* squared change = .058, *F* change (1, 77) = 5.19, p = .026). The final model was significant and explained 13.8%, *F* (4, 77) = 3.09, *p* = .02 of variance, with gender as the only significant independent predictor (*beta* = .24, p = .026).

Gender also contributed a significant 4.4% to PLS scores at 24 months after controlling for UK-CDI scores at 18 months (*R* squared change = .044, *F* change (1, 78) = 5.46, p = .022). The final model explained 37.3%, *F* (3, 78) = 15.46, *p* < .001. In this situation, both Comprehension at 18 months (*beta* = .42, p = .001) and gender (*beta* = .21, p = .022) contributed significantly in the final model compared to the previous model in which only gender but not UK-CDI scores at 12 months could add significant variance to PLS scores at 24 months. If entered as the first step gender contributed 7.1% of variance to the full model and showed the same beta and p values as if entered as a second step.

Furthermore, it was examined if girls had better language than boys between 12 and 36 months. In the current sample, a one-way between- groups MANOVA was performed to investigate sex (IV) differences in communication and language scores (DV, i.e. all 16 variables between 12 and 36 months). The model was significant (*F* (16, 73) = 2.14, *p* = .015; Wilk's Λ = .681, partial η^2 = .32.). Girls (*M* = 47.91, *SD* = 7.33) were better than boys (*M* = 42.41, *SD* = 8.20) at using UK-CDI Gestures at 18 months (*F* (1, 88) = 10.46; *p* = .002; partial η^2 = .11). In addition, girls (at 18 months: *M* = 108.43, *SD* = 12.24; at 24 months: *M* = 106.34, *SD* = 10.71) had significantly higher scores than boys (at 18 months: *M* = 102.04, *SD* = 11.04; at 24 months: *M* = 100.09, *SD* = 11.16) in terms of general language ability (PLS) at 18 months (*F* (1, 88) = 6.58; *p* = .012; partial η^2 = .07) and at 24 months (*F* (1, 88) = 6.92; *p* = .010; partial η^2 = .07) and the effect sizes were moderate. However, girls and boys did not differ significantly in terms of language scores at 36 months. As there was a high variation in age of the participants at the early testing points, two MANCOVAs were employed separately for communication scores at 12 and 18 months to control for age (age at 12 or 18 months as covariate) and the results remained the same. Overall, the results show that on average girls and boys achieved scores within the normal range for Gestures (Median: 44) and the PLS instrument (Mean: 100, normal range: 85-115) but within this normal range girls yielded higher scores than boys.

7.2.3.3. Contributions of SES to later language scores

The same hierarchical regressions were used as above in 7.2.3.1., but as a last step maternal education / household income rather than prematurity was entered. Maternal education did not add any significant contributions to later language scores after controlling for UK-CDI scores at 12 or 18 months. Whilst household income added significant variance after controlling for UK-CDI scores at 12 months for one outcome variable (Production 18 months), income did not contribute significantly beyond UK-CDI scores at 18 months. For Production scores at 18 months as the outcome variable, age at 12 months entered as the first step explained .02% variance, Production and Gestures at 12 months entered as the second step contributed 39.4% of variance and income entered as the last step contributed an additional 3% (and 3.5% if entered as the first step), R squared change = .030, F change (1, 77) = 4.04, p = .048. The total variance explained by the model was 42.4%, F (4, 77) = 14.18, $p < .001$ and all predictors were significant in the final model, recording the highest beta value for Production at 12 months (*beta* = .53, p < .001), followed by Gestures (*beta* = .26, p = .016), age (*beta* = -.25, p = .015) and income (*beta* = .18, p = .048). Even though income added significant variance to the model, the amount of variance explained was a lot less than by the UK-CDI scores. This was confirmed by another regression analysis in which income was entered as the first step and contributed 3.5% of variance.

Table 35. *Production scores between 12 - 36 months separate for household incomes**

	£0 to £14000		£14001 to £24000		£24001 to £42000		£42001 or more	
	N	M (SD)	N	M (SD)	N	M (SD)	N	M (SD)
Production 12m	2	10 (4.24)	8	4.75 (3.96)	34	9.35 (14.97)	34	8.5 (16.59)
Production 18m	2	19.5 (2.12)	8	49 (60.03)	34	65.97 (80.96)	34	65.94 (49.23)
Production 24m	2	354 (371.94)	8	226.13 (165.92)	34	280.97 (183.56)	34	325 (143.57)
Production 36m	2	50 (70.71)	8	72.5 (26.68)	34	66.18 (28.09)	34	77.91 (19.09)

*Data was only reported of two-adult households thus the data for children of three single parents was not reported here

It should be also considered that the groups were not equally distributed, see Table 35 below. When collapsing the two lowest groups, it becomes clear that children with parents of lower incomes reported smaller productive vocabularies on average between 12 and 24 months than those parents in the two higher income brackets.

The literature suggests that for younger children parents with high SES report smaller vocabularies for their children than parents with low SES. In contrast, older children from higher SES backgrounds showed better language ability on the CDI compared to children from low SES families (e.g., Fenson et al., 1994). We used MANOVA analyses to assess if SES (maternal education or household income used as IV) had a significant effect on language scores collected between 12 and 36 months (DVs). Maternal education was measured on a nominal scale with 5 levels and household income had four levels (see Analysis 1 Table 16). If the models were significant, post-hoc tests would establish if the same pattern was present as reported in the literature. However, the model was neither significant for maternal education (F (64, 210) = 1.11, p = .290; Wilk's Λ = .319, partial η^2 = .25) nor for household income (F (48, 161) = 1.29, p = .122; Wilk's Λ = .380, partial η^2 = .28).

Furthermore, we assessed if there was a significant interaction between time and the two highest household income groups (£24001 to £42000, £42001 and more), but the interaction was not significant, Wilk's Lambda = .96, F (3, 64) = .11, p = .95, partial eta squared = .01. The main effects for time (Wilk's Lambda = .99, F (3, 64) = .26, p = .86, partial eta squared = .01)[45] and household income (F (1, 66) = .1.16, p = .29, partial eta squared = .02) were also not significant, suggesting no difference in Production scores of the two household income groups.

As the participants were distributed unevenly across the levels with a bias towards high SES, the maternal education data was changed into a dichotomous format to compare low (N = 28) and high

[45] Z-scores were used for this calculation, hence no significant main effects for time were expected.

(N = 54) SES groups using the same definition as Fenson et al. (2007). However, the model was still not significant (16, 56) = .935, p = .536; Wilk's Λ = .789, partial η^2 = .21).

The effect of SES on language scores was not significant in the MANOVA analysis and only one small effect was found amongst several regression analyses. This could be because the current sample was biased towards higher SES backgrounds. In total, there were only 34% of children from low SES backgrounds (as defined by Fenson et al., 2007) and out of those most children had mothers with A-Level education (75%), some mothers had GCSE education (21%) and only one mother had no formal education (4%). It should be noted that this type of research often has problems with recruiting low SES families. Possibly due to the bias towards high SES backgrounds, no significant SES differences could be found in the current sample, which does not mean that differences in SES do not exist.

7.2.3.4. Contributions of sibling status to later language scores

The same hierarchical regressions were used as above in 7.2.3.1., but as a last step sibling status rather than prematurity was entered. Sibling status added a significant 4.3% of variance to Production scores at 24 months after controlling for UK-CDI scores at 12 months (R squared change = .043, F change (1, 77) = 4.08, p = .047). The final model explained 18.3%, F (4, 77) = 4.31, p = .003 in which two variables were significant. Sibling status was significant (*beta* = -.21, p = .047) but Production at 12 months had a higher beta value (*beta* = .31, p = .009). Furthermore, upon entering sibling status as the first step, the variable only contributed 1.9% of variance in contrast to 4.3% if entered after as the last step. In contrast to gender, sibling status was not a very robust predictor for language at 24 months as it did not remain significant after controlling for UK-CDI scores at 18 months. The data suggests that children with older siblings produce more words at 24 months compared to first-born children, see Table 36 below. This finding is somewhat unexpected and possible reasons for this result will be given in the discussion.

Table 36. *Production raw scores between 12-36m separate for sibling status*

	First child		Child with older siblings	
	N	M (SD)	N	M (SD)
Production 12m	50	10.28 (17.74)	32	5.34 (5.63)
Production 18m	50	70.16 (75.16)	32	52.03 (42.76)
Production 24m	50	284.62 (167.56)	32	317.13 (172.99)
Production 36m	50	71.64 (25.39)	32	71.69 (27.56)

A repeated measures ANOVA investigated a possible interaction effect for Production scores between 12 and 24 months (DV) in terms of sibling status (IV). The main effect of time was not significant (Wilk's Lambda = 1.0, F (2,79) = .12, p = .89, partial eta squared = .003) but the interaction between time and sibling status was significant (Wilk's Lambda = 9.3, F (2,79) = 3.19, p = .046, partial eta squared = .08). Even though, this suggests a moderate effect, the pairwise comparison table showed no significant differences after correcting for Bonferroni correction. Similarly, a repeated measure ANOVA with all four Production scores was also not significant in terms of time (Wilk's Lambda = .99, F (3,78) = .16, p = .92, partial eta squared = .01) or the interaction between time and sibling status (Wilk's Lambda = .92, F (3,78) = 2.31, p = .08, partial eta squared = .08).

Furthermore, a MANOVA analysed the effects of sibling status (IV) on language scores between 12 and 36 months (DVs) and the model was significant and demonstrating a large effect (F (16, 85) = 2.09, p = .016; Wilk's Λ = .718, partial η^2 = .28) after weighing the means to balance the unequal sample size. However, the subsequent ANOVAs showed no significant communication differences in terms of sibling status. There was more variation in age at 12 and 18 months, therefore the same MANOVAs were run again with age as the covariate. After controlling for age, the models were neither significant at 12 months (F (4,96) = .70, p = .594; Wilk's Λ = .972, partial η^2 = .03) nor at 18 months (F (5,95) = .70, p = .655; Wilk's Λ = .966, partial η^2 = .03).

It was further assessed if the number of siblings (between 0 - 4 or more) significantly correlated with communication scores. As sibling status was not normally distributed according to the Kolmogorov-Smirnov test, a Spearman's rho test was used. A positive and small correlation was found between number of siblings and Gestures at 18 months (rs = .266, p = .017), this means children with more siblings used more different types of Gestures at around 18 months. However, after adjusting for Bonferroni correction, this was not significant anymore.

In sum, some analyses found an influence of sibling status on Production and Gestures. However, this should be interpreted with caution (regarding the type 1 error) as several tests were conducted overall and only few found small effects.

7.2.3.5. Ear infections and language development

There were only three children with recurring ear infections in this sample, therefore the analysis of their language development will be descriptive. These three children were very different at the outset of language development, the only girl amongst two boys had average to above average skills in all areas on the CDI at 12 and 18 months. This was also reflected by the PLS results on which she achieved above average skills (two SDs above the mean) at 18 months and skills within the normal range but above average at 24 months (PLS score = 107). Child 2 achieved just below median scores at 12 months but his scores on the CDI were well behind 18 months in terms of vocabulary (below 5th percentile) and gestures (25th percentile). His PLS score was also low as he achieved a standard score of 85 which is the lowest score within the normal range. However, at 24 months his score on the PLS was average (PLS score = 99). Child 3 showed scores between the 10th and 25th percentile at 12 months and scores below the 10th percentile on all areas of the CDI at 18 months. His PLS score was also below the normal range, between 1 - 2 SDs from the mean at 18 months (PLS score = 81) and at 24 months (PLS score = 79). Interestingly, at 36 months all children had scores within the normal range according to the ASQ questionnaire. Their scores were also not below their peers in terms of the 3-year language measure for which no norms exist at presence.

7.2.3.6. Contributions of sleep to later language scores

The same hierarchical regressions were used as above in 7.2.3.1., but as a last step sleep at 12 months / at 18 months rather than prematurity was entered. Sleep at 12 months did not add any significant contributions to later language scores after controlling for UK-CDI scores at 12 months or at 18 months. However, sleep at 18 months nearly contributed significant variance to Production 24 months scores beyond UK-CDI at 18 months (R squared change = .024, F change (1, 78) = 3.91, p = .052). The final model explained 51.6%, F (3, 78) = 27.72, p < .001, and Production at 18 months had the highest beta value (*beta* = .41, p < .001) followed by Comprehension at 18 months (*beta* = .32, p = .004) and a near significant result of sleep at 18 months (*beta* = -.16, p = .052).

As hours of sleep were not normally distributed, the following analyses used non-parametric tests. When children were aged around 12 months (*Mean* = 12.99, *SD* = 1.30) parents reported significantly more sleep than at 18 months (*Mean* = 12.73, *SD* = 1.13) according to a Wilcoxon signed-rank test (Z = -2.3, p = .044), see Appendix 14.

Furthermore, hours of sleep at 12 months did not correlate with any of the concurrent or subsequent language variables. However, at hours of sleep at 18 months correlated negatively but weakly with PLS Expressive Communication scores at 18 months (rs = -.314, p = .006) and later Comprehension (rs = -.266, p = .019) and Production (rs = -.226, p = .046) at 24 months; however, after controlling for Type 1 error using the Bonferroni correction the correlations were not significant anymore.

It was further investigated if sleep at 18 months served as a moderator between language skills at 18 months (PLS Expressive Communication, PLS EC) and Production and Comprehension at 24 months. There were no long-term moderating effects of sleep (at 18 months) between early and later communication skills, as the interaction between sleep and PLS EC was not significant (p = .959) with Production 24 months as the outcome variable, neither was the interaction between sleep and PLS EC (p = .476) with the outcome variable of Comprehension at 24 months.

The analyses here indicate that there might be a relationship between sleep and language that is worth investigating further. Even though from the current data we cannot say that sleep significantly influences language development, if examined in a different context (for example the influence of hours and quality of sleep on language) may give an opportunity to examine possible moderating or mediating effects.

7.2.3.7. Contributions of family risk of dyslexia or speech or language problems to later language scores

The same hierarchical regressions were used as above in 7.2.3.1., but as a last step family risk status rather than prematurity was entered. Whilst at-risk status (dichotomous variable) could not explain any additional variance in later language scores after controlling for UK-CDI scores at 12 months, it contributed some additional 5.4% of variance (3.9% if entered as first step) to grammar scores (Sentence Complexity) at 36 months beyond UK-CDI scores at 18 months (R squared change = .054, F change (1, 78) = 5.93, p = .017). This means the total model explained 29.3%, F (3, 78) = 10.78, $p < .001$ in which only Production at 18 months (*beta* = .39, p = .003) and family risk status (*beta* = -.23, p = .017) were significant.

In addition, a one-way MANOVA was used to explore if children with familial risk performed worse on language tests compared to the typical children. Weighted means were used to balance the group sizes for at-risk children who have a first degree relative with dyslexia or a speech or language problem (N = 19) and children with no familial risk (N= 63). The independent variable was familial risk and the dependent variable was communication scores between 12 and 36 months. There was a statistically significant difference in communication scores depending on familial risk, F (16, 73) = 2.18, p = .013; Wilk's Λ = .677, partial $\eta^2 = .32$. Parents reported significantly higher Production scores at 24 months for at-risk children (M = 350.82, SD = 216.73) compared to typical children but the effect size was small (M = 273.46, SD = 153.50), (F (1, 88) = 4.43; p = .038; partial $\eta^2 = .048$). This finding is unexpected and possible reasons for this will be given in the discussion.

Furthermore, it should be noted here that Sentence Complexity as an indicator of grammatical development did not differ across groups at the same age (F (1, 88) = .297; p = .587; partial η^2 = .003). In a MANCOVA using age at 24 months as covariate with the communication variables at 24 months as dependent variables, the model was still significant (F (4, 84) = 2.85, p = .029; Wilk's Λ = .880, partial η^2 = .12) and the effect of family risk on production scores remained the only significant relationship at this age but showing a small effect size (F (1, 87) = 4.26; p = .042; partial η^2 = .047).

A one-way ANOVA was used with at-risk status (IV) and production at 12 months (DV). The difference in production scores at 12 months for at-risk children (M = 4.29, SD = 5.15) and typical children (M = 7.43, SD = 11.51) nearly reached significance (F (1, 88) = 3.92; p = .051; partial η^2 = .043) with the inverse relationship showing lower scores in at-risk children. The same pattern was found for Sentence Complexity scores at 36 months for at-risk children (M = 6.94, SD = 4.62) and children without risk (M = 8.55, SD = 3.66), this also nearly reached significance (F (1, 88) = 3.90; p = .052; partial η^2 = .042).

As we found that there was a high variation in age at 12 months, the next step was to include age at 12 months as a covariate in the one-way MANOVA to investigate the effect of familial risk (IV) on all UK-CDI variables at 12 months (DVs) after controlling for age. The overall model was significant (F (4, 84) = 2.95, p = .025; Wilk's Λ = .877, partial η^2 = .12). After controlling for age at 12 months, Production reached significance (F (1, 87) = 3.95; p = .050; partial η^2 = .043) which means that at-risk children produced significantly fewer words than children without risk at around 12 months of age but the effect size was small.

Overall, at-risk children showed lower vocabularies during the very early stages of language development and lower grammatical ability than their typical peers at a time when grammar is established. It is interesting that parents of at-risk children compared to typical children reported significantly higher vocabularies but no differences in grammatical ability at 24 months- a time when the use of an array of words is common but the first complex grammatical structures are still emerging. It is possible that parents of at-risk children complete questionnaires in a different way compared to typical parents, particularly when faced with the longest vocabulary checklist used in this study, the Lincoln Toddler CDI at 24 months. This would make even more sense if the at-risk child's parent who completed the survey was affected by dyslexia or a language deficit themselves. However, further investigation is required to confirm these assumptions. Nevertheless, as the same parents indicate lower abilities in early vocabulary and later grammar, the UK-CDI and the 3-year language instrument are in line with previous research findings.

7.2.3.8. Contributions of cognitive skills to later language scores

Cognitive skills were not collected prior to 18 months, thus it was only examined if cognitive skills could contribute additional variance beyond UK-CDI scores at 18 months. The same hierarchical regressions were used as above in 7.2.3.1., but as a last step cognitive skills at 18 months rather than prematurity were entered. Indeed, cognitive scores explained an additional 8.9% of variance in PLS scores six months later (R squared change = .089, F change (1, 78) = 11.97, p = .001). This final model explained 41.8%, F (3, 78) = 18.69, p < .001. Interestingly, cognitive skills had a higher beta value (*beta* = .32, p = .001) than Production (*beta* = .28, p = .019) and Comprehension (*beta* = .27, p = .033). If entered in step 1, cognitive skills explained a significant 17.3% of variance (*beta* = .32, p = .001).

The next analysis investigated if children's cognitive scores (standard scores) were correlated between 18 and 24 months. As cognitive scores were not normally distributed according to the Kolmogorov-Smirnov test, a non-parametric Spearman's rho correlation was used. The relationship between children's cognitive scores at 18 and 24 months was moderate, positive and statistically significant (rs = .59, p < .001). There were several significant associations between communication scores and cognitive scores. Even after adjusting for Bonferroni correction (p (0.05/16) = .003), especially cognitive scores at 24 months showed several significant correlations between CDI scores (Gestures at 12 months, Comprehension at 18 and 24 months, Sentence Complexity at 36 months) and general language ability (PLS) at 18 and 24 months. The direction of relationship was always positive, and the strength was weak to moderate, see Table 37.

Table 37. *Spearman's rho correlations between cognitive skills and communication abilities*

	cognitive skills 18m			cognitive skills 24m		
	Correlation Coefficient	Sig. (2-tailed)	N	Correlation Coefficient	Sig. (2-tailed)	N
Comprehension 12m	.226*	.045	79	.175	.123	79
Production 12m	-.004	.971	79	.143	.207	79
Gestures 12m	.259*	.021	79	.392***	.000	79
Phrases Understood 12m	.138	.225	79	.177	.119	79
Comprehension 18m	.314**	.005	79	.354***	.001	79
Production 18m	.115	.312	79	.204	.071	79
Gestures 18m	.224*	.050	77	.308**	.006	77
Phrases Understood 18m	.237*	.035	79	.288*	.010	79
PLS total 18m	.442***	.000	79	.452***	.000	76
Comprehension 24m	.315**	.005	79	.385***	.000	79
Production 24m	.270*	.016	79	.323**	.004	79
Sentence Complexity 24m	.210	.064	79	.282*	.012	79
PLS total 24m	.484***	.000	75	.534***	.000	78
Production 36m	.193	.089	79	.273*	.015	79
Sentence Complexity 36m	.204	.071	79	.385***	.000	79
ASQ communication 36m	.148	.192	79	.221	.051	79

***. Correlation is significant at the 0.003 level (2-tailed).

**. Correlation is significant at the 0.01 level (2-tailed).

*. Correlation is significant at the 0.05 level (2-tailed).

It was further investigated if cognitive skills at 18 months had a moderating effect on the relationship between earlier and later language skills but the interaction was not significant for any of the analyses, see Appendix 15.

Non-parametric partial correlations were also run for communication variables at 12 and 18 months and age at 12 and 18 months as the control variables. Generally, the strength and directions of the relationships remained the same after controlling for age.

In conclusion, it is interesting that cognitive skills at 18 months added almost as much variance as UK-CDI scores at 18 months to PLS scores at 24 months, but that this was not significant for any other language and grammar outcome variables at 24 and 36 months. As cognitive skills explained so much variance in later PLS scores, it is possible that the PLS might be confounded by cognitive ability. It may also suggest that the individual variation in standardised test results is driven by children's ability to perform in testing situations. Cognitive scores at 18 months were correlated with general language ability (PLS) at 18 and 24 months but at this age cognitive scores showed no significant associations with concurrent and subsequent vocabulary and grammar. It seems that the progress in cognitive development between 18 and 24 months was important to find links with grammar ability at 36 months. This can be explained by the significant but only moderately strong relationship between cognitive skills at 18 and 24 months. Previous research has shown that cognitive and language development are linked (Rose et al., 2009); however, as already described for correlations between language scores at different ages point (see Chapter 5), relationships were usually stronger if the gap between testing times was shorter. Table 37 shows that whilst cognitive skills at 18 months were not related to grammar at 36 months, a positive and moderately strong association was found between cognitive skills 24 months and grammar at 36 months. Overall, cognitive ability was consistently associated with vocabulary comprehension and general language ability (PLS).

7.2.3.9 Contributions of motor skills to later language scores

Motor skills were collected at 18 and 24 months using the Bayley scales. Therefore, it was assessed if motor skills at 18 months could add significant variance after controlling for CDI scores at 18 months. The same hierarchical regressions were used as above in 7.2.3.1., but as a last step motor skills at 18 months rather than prematurity were entered. Motor ability at 18 months added another 8.3% of variance (and even 20.4% if entered as first step) to PLS scores at 24 months after UK-CDI scores at 18 months had been entered in the model (R squared change = .083, F change (1, 78) = 10.94, p = .001). The full model explained 41.2% of the variance in PLS scores at 24 months F (3, 78) = 18.19, p < .001, and Comprehension had a somewhat higher beta value (*beta* = .31, p =

.012) than motor skills (*beta* = .30, p = .001) and Production did not contribute significantly to the final model (*beta* = .21, p = .075).

Further investigations examined if motor skills at 18 and 24 months were associated with communication scores between 12 and 36 months. First, it was established that motor scores between 18 and 24 months correlated positively and strongly and the relationship was significant (rs = .66, p < .001). Table 38 below shows the associations; however, after using the new cut-off (p = .003) motor skills at 18 months correlated significantly with general language (PLS) at 18 and 24 months. Furthermore, motor skills at 24 months were also associated with general language ability at 24 months but also with measures of CDI language comprehension at 18 months (Comprehension and Phrases Understood). All relationships were positive and weak to moderate in strengths. It should also be mentioned that motor and general cognitive skills (Bayley test scores) correlated significantly and moderately at 18 months (rs = .48) and 24 months (rs = .54).

Table 38. *Spearman's rho correlations between motor skills and communication abilities*

	motor 18m			motor 24m		
	Correlation Coefficient	Sig. (2-tailed)	N	Correlation Coefficient	Sig. (2-tailed)	N
Comprehension 12m	.026	.820	78	.098	.395	78
Production 12m	-.078	.496	78	.055	.634	78
Gestures 12m	.157	.170	78	.317**	.005	78
Phrases Understood 12m	.135	.239	78	.296**	.009	78
Comprehension 18m	.312**	.005	78	.361***	.001	78
Production 18m	.241*	.034	78	.252*	.026	78
Gestures 18m	.313**	.006	76	.321**	.005	76
Phrases Understood 18m	.269*	.017	78	.352***	.002	78
PLS Total Language 18m	.387***	.000	78	.296*	.010	75
Comprehension 24m	.267*	.018	78	.237*	.037	78
Production 24m	.283*	.012	78	.274*	.015	78
Sentence Complexity 24m	.210	.065	78	.129	.260	78
PLS Total Language 24m	.463***	.000	74	.502***	.000	77
Production 36m	.275*	.015	78	.207	.069	78
Sentence Complexity 36m	.213	.061	78	.234*	.039	78
ASQ Communication 36m	.254*	.025	78	.199	.081	78

***. Correlation is significant at the 0.003 level (2-tailed).

**. Correlation is significant at the 0.01 level (2-tailed).

*. Correlation is significant at the 0.05 level (2-tailed).

A more detailed investigation of the PLS scores revealed that the effect was carried more by the Auditory Comprehension subscale rather than the expressive communication scale particularly for PLS scores at 18 months, see Table 39.

Table 39. *Spearman's rho correlation for separate PLS subscales and motor skills*

	motor 18m			*motor 24m*		
	Correlation Coefficient	Sig. (2-tailed)	N	Correlation Coefficient	Sig. (2-tailed)	N
PLS Auditory Comprehension 18m	.448**	.000	78	.320**	.005	75
PLS Expressive Communication 18m	.239*	.035	78	.195	.093	75
PLS Auditory Comprehension 24m	.489**	.000	74	.576**	.000	77
PLS expressive Communication 24m	.334**	.004	74	.325**	.004	77

**. Correlation is significant at the 0.01 level (2-tailed).

*. Correlation is significant at the 0.05 level (2-tailed).

The next step was to test for moderating effects of motor skills at 18 months between earlier and later language skills but the interaction was not significant for any of the analyses, see Appendix 15.

After using partial correlations to control for age at 12 and 18 months, the strengths of correlations remained very similar and the direction of relationships remained positive. In contrast to cognitive skills which only correlated moderately between 18 and 24 months, motor skills were associated strongly. This strong positive linear relationship between 18 and 24 months for motor skills can explain why non-significant relationships between motor scores at 18 months and language scores at 36 months continued to be non-significant between motor scores at 24 months and language at 36 months. In sum, motor ability at 18 and 24 months was significantly correlated with communication variables at 18 and 24 months, particularly for comprehension. Therefore, it is possible that the relationship may remain significant with later language comprehension, but this was not tested in the current study.

7.3. DISCUSSION

The aim of this investigation was to establish if UK-CDI scores at 12 months could explain any more variance beyond UK-CDI scores at 18 months. It was further analysed which other factors contributed to language development and if they could explain additional variance beyond the UK-CDI scores.

Overall, earlier language scores explain the most variance in later language ability and thus are the best single predictor for future outcomes. Even though other factors also added significant variance, the strength of their contributions varied depending on outcome measure used and time tested.

UK-CDI scores obtained at around 12 months could not explain any additional variance beyond UK-CDI scores at 18 months, this is important as it may have clinical implications. For example, in Germany children's communication abilities are documented at around their first birthday and again later during their development (Gemeinsamer Bundesausschuss, 2016). This early investigation is useful as it may help to detect other underlying problems (e.g. hearing deficits) which occur at the same time. However, for making predictions about future language ability, the CDI's additional use at around 12 months is limited. Although it appears that the UK-CDI at 12 months cannot give any more additional information as obtained by the UK-CDI at 18 months, the previous analyses showed that for the lowest Production ability subgroup at around 12 months, grammar scores at 36 months were a lot lower than grammar scores for the lowest Production group at 18 months. This shows that particularly for the lowest ability children, scores at around 12 months rather than 18 months could be important to make predictions about future grammar skills. Future research should investigate if children with very low Production scores at around 12 months and grammar scores (i.e. Sentence Complexity) at 3 years differ from other children on a range of different language areas beyond three years of age.

Overall, these findings suggest that the UK-CDI only needs to be administered once at around 18 months to cover the majority of children and does not require a second earlier testing as this would increase the cost to the health system and would not profoundly improve the ability to predict children's future ability. While some research suggested that CDI-type questionnaires cannot be recommended for the early identification of developmental language disorders at 12 months due to insufficient predictive validity at this age (Sachse, Saracino et al., 2007), the current results suggest that there is already some stability at 12 months but the UK-CDI at 18 months is able to explain more variance in later language scores.

Furthermore, family, biological and environmental influences were able to explain some additional variance beyond the UK-CDI. Their contributions were not consistent across all later language categories as one predictor never explained additional variance in more than one outcome variable. If significant additional variance was found, it usually did not contribute higher beta values than did the CDI scores, except for gender after controlling for CDI variables at 12 months and cognitive scores after controlling for CDI scores at 18 months.

Overall, the highest additional variance was contributed by cognitive skills, supporting research which demonstrated that infant information processing skills significantly predicted language skills at 36 months beyond Infant CDI scores (Rose et al., 2009). The most consistently reported factors in the literature to improve screening results for language outcomes were family history of speech and language delay, male gender, parent educational level as well as perinatal influences (Reilly et al., 2007). Indeed, we also found significant contributions which were of similar strength as reported in previous research for family risk of dyslexia or speech or language problems, male gender and income but also sibling status which was mentioned less consistently in the literature (Reilly et al., 2007; Rescorla, 2011).

In addition, we found that at two years of age, CDI Production scores were higher for children with older siblings (vs. children with no older siblings) and those children with familial risk for language and literacy related problems (vs. children with no familial risk). This an unexpected finding as previous research repeatedly reported slower language development for at-risk children (e.g., H. Lyytinen et al., 2001; P. Lyytinen et al., 2001) and children with older siblings (e.g., Berglund et al., 2005; Hoff-Ginsberg, 1998; Pine, 1995). In our data, parents of these children reported significantly lower scores at younger ages and at 3 years the abilities were of similar or expected direction again. It needs to be investigated if the Lincoln Toddler CDI led parents to complete the questionnaire with less care due to its length (689 words). This may have been the case particularly for those parents with less time at home due to more children as part of the household or some parents who had difficulty filling in the form due to dyslexia. This assumption is plausible as these children did not differ in terms of PLS scores at the same age when children were assessed by an examiner. However, it should be mentioned that only one parent mentioned that she had difficulty completing the Toddler CDI due to her dyslexia. None of the other parents gave any feedback or had complaints about any of the parent-report instruments.

Previous research indicates that the contributions of environmental factors were stronger for language development during early childhood (2 – 4 years) than genetic influences and the strength of contributions reversed during middle childhood when genetic influences became more influential for language (Hayiou-Thomas et al., 2012). Our data also showed that some environmental factors (sibling status and income) and biological influences which are mediated by the environment (gender and familial risk status) contributed to later language scores. In contrast to Hayious's study (2012) which thoroughly examined these influences in terms of zygosity status (identical versus fraternal twins), we used observable variables such as income and sibling status across the children which may be less sensitive. We found that the environmental and biological contributions were all small and explained at most 5% of variance, similarly to previous research (e.g., Berglund et al., 2005; Fenson et al., 1994).

It seems that developing skills (e.g. cognitive or motor skills) were better able than more inert biological (e.g. gender) or environmental (e.g. sibling status) influences to contribute to later language development. Inert influences contribute only small variances to later development. We know that cognitive or motor skills are also shaped by different biological and environmental influences (e.g., Kahan et al., 2017; Tucker-Drob & Briley, 2014). The variability between children can be documented better on standardised tests (for cognitive and motor skills) than for categories like gender for which children can fall only into one or the other category. Furthermore, these emerging cognitive and motor skills are better predictors than inert influences as they reflect child development in a more flexible and representative way. This could be explained in the context of the neuroconstructivist approach to language development (D'Souza, D'Souza, & Karmiloff-Smith, 2017; Mareschal et al., 2007). This theory explains that child development happens in adaptive systems which are complex as they rely on the interdependence with related cognitive domains (memory, attention) and other modalities (e.g. vision, hearing, motor functions) and contexts (i.e. genes and social interaction) across time. This means a child with less favourable perinatal factors and little stimulation is likely to show lower cognitive and motor skills and later lower language skills than their peers with average perinatal outcomes and stimulation.

Some factors (e.g. gender, motor) had an influence on language at 18 and 24 months but not at 36 months. This could be due to the instruments used at 36 months. For example, the ASQ measures a range of communicative abilities has been constructed for distinguishing between delayed and typical children, thus differentiation between abilities on all levels was not the aim and possible group differences may not be visible due to the design of this instrument. On the other hand, vocabulary has been assessed throughout the study and general language has been assessed twice using the PLS instrument; however, even when using the same measurements biological and environmental influences could not be found consistently. We detected most differences at 24 months. It is possible that biological and environmental influences are important for language development in children younger than 2 years of age, but they may not yet be easily detectable in language tests. In contrast to these biological or environmental influences, other developing abilities (e.g. information processing or gestures) show more variability with regards to test scores. It is possible that these abilities (e.g. gestures) serve as mediators between biological or environmental influences (e.g. parental communication) and later language (e.g., Rowe et al., 2008) and that the correlations between these other developing skills (e.g. gestures) and later language are stronger than the direct effect between biological/environmental factors and later language, and thus are better at contributing more variance to later language scores. In addition, the 3-year language measure which is a brief vocabulary and grammar checklist similar to the CDI questionnaires employed between 12 and 24 months was also used at 36 months. At around 12 months the UK-CDI almost fully reflects the child's language and gestures; however, as children get older their language becomes too complex to document every sentence they utter. Therefore, the selection of items which distinguish between children and across age becomes even more important for the design of the instrument. It may be due to item selection or the language area assessed that we could not detect any subtle differences that may exist between groups. It is also possible that the influence of biological and environmental factors is associated with language at differing strengths depending on the stage of development. For example, in neural network modeling it is a standard principle of learning that biological and environmental factors have more influence in early stages of learning in a network - which does not mean that these other factors may not also have an influence on development along the way. In later stages of learning, the content that the network has already learnt plays a bigger part in the quality of the learning (Elman et al., 1996). This approach coincides with the current findings that biological and environmental factors play a less important role than early language ability in predicting later language development.

As the contributions of additional factors other than early language on later language ability were inconsistent in strengths as they varied depending on outcome measures and time tested, the decision was made to rely on UK-CDI scores for making predictions about language delay on the individual level. The next chapter investigates if the UK-CDI can make meaningful predictions for individual children about future language delay using different definitions of delay. If language delay can be predicted, specific cut-off scores on the UK-CDI are suggested.

8. ANALYSIS 4: INDIVIDUAL-LEVEL ANALYSES FOR PREDICTING LATER LANGUAGE STATUS

The question to answer in this analysis is whether UK-CDI scores can predict later language status (delay vs no delay) at an individual level rather at group level. More specifically, it was examined if UK-CDI scores at around 12 or 18 months could reliably predict language status at 24 or 36 months. Whilst there is little agreement about the most effective timing for detecting language problems and type of intervention (Law, Garrett, & Nye, 2004), there is some evidence that language delay can be detected at around 24 months (e.g., Rescorla, 2013; Sachse et al., 2007; Ullrich & von Suchodoletz, 2011). Some studies also included children between 18 and 35 months of age for studying expressive language delay (late-talking) which is the most common occurring type of delay (Rescorla, 2011). If the newly developed UK-CDI used at around 12 or 18 months can correctly predict language delay for individual children at 24 or 36 months, this could support its use for clinical purposes as a screening tool. This analysis will also find out if children who were delayed at 24 months remain delayed 12 months later.

8.1. METHOD

8.1.1. Participants

The number of participants were the same as in the studies above (see Chapter 5 under 5.1.1. Participants).

8.1.2. Materials

The same instruments were used as in the previous studies.

8.1.3. Procedure

The procedure was the same as in the previous studies.

8.1.4. Statistical analysis

The first step was to evaluate which criteria for delay[46] should be used at 24 and 36 months.

8.1.4.1. Criteria for language delay at 24 months

At 24 months, the criterion for late-talking language delay is a performance below the 10th percentile on the CDI (Dale, Simonoff, Bishop, & Plomin, 1998; Ellis Weismer, 2007). Originally, a criterion of an expressive vocabulary (Production) of less than 50 words or a lack of two-word combinations was proposed for the Language Development Survey (LDS) questionnaire (Rescorla, 1989) which was later adapted to other CDI-type questionnaires in German (Ullrich & von Suchodoletz, 2011). A general language delay can be detected using the Preschool Language Scale (PLS) with a cut-off for delay from 1 standard deviation below the mean (Zimmerman et al., 2014).

8.1.4.2. Criteria for language delay at 36 months

At 36 months, we used the British ASQ which has norms available for American children and those are commonly used for British children (Bedford et al., 2013). Children with a communication deficit are detected if their scores fall below two standard deviations from the mean (Squires et al., 2009). Cut-offs are not available for the 3-year parent report language measure (Dionne et al., 2003) as norm data does not exist yet.

---

[46] As there are no norms for either the Lincoln Toddler CDI, the British ASQ or the 3-year parent-report measure we adapt cut-offs from questionnaires in other languages. Thus, we cannot be entirely certain that the same cut-offs apply to British children. Nevertheless, for ease of reading, I will use the term *delay* from now on for all children who fall into the *criteria for language delay categories* according to my specification.

8.1.4.3. Analysis using the criteria for language delay

Receiver operating characteristic (ROC) curve analyses were used to predict language status at 24 and 36 months using the previously selected criteria for language delay. The classification accuracy (specificity, sensitivity) and likelihood ratios (positive likelihood ratio = sensitivity/ 1 - specificity, negative likelihood ratio = 1 – sensitivity/ specificity) were calculated. These ROC analyses were suitable to detect which cut-off scores on the UK-CDI were most effective to make clinically meaningful predictions separately for 24 and 36 months of age. ROC-curves which showed at least 80% specificity and sensitivity are regarded to be useful for clinical purposes as suggested by Westerlund et al. (2006). It was further investigated how many children remained delayed between 24 and 36 months. In analyses with missing data, participants were removed; however, their data was used in other analyses if the required data was available.

8.2. RESULTS

8.2.1. Determining the criteria for language delay at 24 months

There were 78 children who took part in the PLS assessment at 24 months. Their results are compared with their CDI ability status at the same age. There were five children with a general delay in terms of the PLS at 24 months. For three children the delay was mainly carried by both domains (Expressive Communication and Auditory Comprehension) and for two children the delay was mainly depending on the Auditory Comprehension subscale of the PLS but their expressive skills were also below the mean.

When comparing those five children to the traditional late-talking diagnostic criteria by Rescorla (1989) only two children were delayed on both tests. These two children had low PLS scores in both subscales and on the CDI, one of them had a low number of words but word combinations (20 words) and the other had sufficient words (65 words) but no word combinations. In addition, using Rescorla’s criterion we identified another 9 children with a delay on the CDI. Out of the 11 children with a CDI delay, there were children with low Production (expressive vocabularies) and no two-word combinations (N=3), only grammar (no two-word combinations) delays (N=7) and only Production delays (N=1).

Out of those 78 children in total, there were 3 children with a severe delay and 6 children with a mild delay on the ASQ at 36 months as well as 7 children with a Production delay and 3 with a grammar delay on the 3-year parent report instrument. Only four children had a delay on both forms at 36 months, the other 11 children had a delay on one form only (total N= 15, 19% of the total).

Combining delay status at 24 and 36 months, there were 14 children delayed in total at 2 years and 9 were still delayed at 3 years (64%), hence 5 children were over-identified (36%). However, out of those 64 with no delay at 24 months, six children developed a delay at 36 months (9%). This means the sensitivity was 60% and the specificity was 92.06%[47]. Sensitivity is the percentage of children who truly have a delay and are correctly identified (true positives) and specificity is the percentage of children who were correctly identified as typically developing (true negatives). Furthermore, the likelihood ratios were calculated which explain the odds that a given result (positive or negative) is expected in an individual who is affected or unaffected by language delay in regards to the instrument used. Tests which are successful at identifying those individuals with a disorder measure have a positive likelihood ratio (LR +) of 10 or more (values of 3 suggest some usefulness) and tests which are efficient at successfully detecting those children without a disorder have negative likelihood ratios (LR -) of .01 or lower; however vales of .30 already suggest some usefulness (Dollaghan, 2013). In our data, the positive likelihood ratio was 7.56 and the negative likelihood ratio was .43 suggesting acceptable ability to identify those children with a delay but limited ability to correctly detect typically developing individuals.

---

[47] Sensitivity is calculated as true positives/ total positives while specificity is calculated as true negatives / total negatives.

The next step was to compare the PLS results with the CDI 10th percentile cut-off (Comprehension = 268.6 words, Production = 70.4 words) at 24 months which may be somewhat less reliable as we do not have any norm data for the Lincoln Toddler CDI. Thus, cut-offs were created from the current data which seemed reasonable as there was little variation in age (Mean = 24.06, SD = .44). We found that there were 7 children with Production delays (of which 5 also had no word combinations) and 7 children with Comprehension scores below the 10th percentile (of which one child also had a grammar delay and two children had a delay in both other categories). Combining those children with Production and Comprehension delays (N=12), 7 children also had a delay at 3 years (58.3%) and 8 children developed a delay later but were not yet identified at 2 years (12%) (sensitivity = 46.67%, specificity = 92.06%). Similarly, if adding those children (N=4) with a lack of word combinations but Comprehension and Production in the typical range at 2 years to the delay status (total N=16), 56% children (N=9) remained in the delay category at 3 years and 9.7% were not recognised by the CDI at 2 years but showed signs of a delay at 3 years (N= 6) (sensitivity = 60%, specificity = 88.89%).

Other authors (e.g., Feldman 2005) chose the 20th percentile (Comprehension = 312.6 words, Production = 115.2 words) cut-off which – when used in this study - varied in prediction performance depending on the combination of delay types, see Table 40. Generally, using these criteria, the number of children (N between 15 and 23) with delays at 2 years increased. This meant a higher correspondence rate with children who were still delayed at 3 years (increased sensitivity) but also a higher number of children over-identified at 2 who did not have a delay at 3 years (decreased specificity). The trend that some children were identified with a delay at 2 years, but then showed no signs of a delay 3 years has been previously described by Feldman et al. (2005).

Table 40. *Number of children with delay, specificity and sensitivity ratings for different combinations of delay if 20th percentile cut-off is used*

	N of delay	sensitivity	specificity	LR +	LR -
Production below 20th percentile	15	56.25%	90.32%	5.81	0.48
Production below 20th percentile AND/OR no word combinations	17	62.50%	88.71%	5.54	0.42
Comprehension AND/OR Production below 20th percentile AND/OR no word combinations	21	62.50%	82.26%	3.52	0.46
Comprehension AND/OR Production below 20th percentile	23	68.75%	80.65%	3.55	0.39

Depending on the criterion for language delay employed here, the percentage of children affected varied between 6.4% and 29.5%. As previous research reports the prevalence of language delay (expressive and mixed delays) was no higher than 20% at 24 months[48] (Law, Boyle, Harris, Harkness, & Nye, 2000b), this could also be considered for choosing the criteria of language delay alongside the previously calculated diagnostic criteria (specificity and sensitivity).

After comparing the long-term results of the Lincoln Toddler CDI at 24 months with outcomes at 3 years, three criteria seem worth further exploring in terms of acceptable specificity and sensitivity levels as well as the percentage of children with a delay. Therefore, the following cut-off levels are used in separate analyses:

1) Less than 50 words or no two-word combinations (following Rescorla's criterion) (N = 11; 14.1%)
2) Vocabulary below 20th percentile AND/OR no word combinations (N = 17; 21.8%) as suggested by Feldman et al. (2005)

[48] Prevalence rates were established using a full literature review which also included many US American studies

3) In addition, PLS criterion (total standard score < 85) as the gold standard for assessing language development (N = 5; 6.4%)

8.2.2. ROC-curve analyses between UK-CDI scores and delay status at 24 months

When using the frequently applied Rescorla criterion of language delay at 24 months (e.g., Rescorla, 1989; Rescorla & Schwartz, 1990; Rescorla, 2011; Rescorla, 2013), none of the CDI variables (Comprehension, Production, Gestures and Phrases Understood) at 12 months were significant on the ROC-curve (see Appendix 16). However, Production at 12 months approached significance ($p = .068$). Using the same criterion but with UK-CDI scores at 18 months, the model was significant for Comprehension at 18 months ($p = .029$) and Production at 18 months ($p < .001$). This means that both vocabulary categories of the UK-CDI at 18 months could significantly discriminate between children with and without a delay at 24 months.

The second criterion (vocabulary below 20th percentile AND/OR no word combinations; e.g., Feldman et al., 2005) includes the same children as counted in Rescorla's criterion as well as an additional 6 children. This time Production at 12 months reached significance ($p = .008$), in contrast to all the other test variables which were not significant. At 18 months, the model was significant for Comprehension ($p = .001$), Production ($p <.001$), Phrases Understood ($p = .021$) but not for Gestures ($p = .145$).

Similarly to Rescorla's criterion, the PLS criterion found that none of the test variables were significant at 12 months but Production was near significance ($p = .066$). In contrast, at 18 months Comprehension ($p = .005$), Production ($p = .006$), Phrases Understood ($p = .013$) and Gestures ($p = .018$) were all significant (see Appendix 16 for an example of ROC-curve analysis).

Importantly, the ROC-curve output can be used to lookup the specificity and sensitivity levels for a range of cut-off scores. If we use the UK-CDI norm data at 12 and 18 months to select cut-off scores for certain percentiles (5th, 10th and 25th percentiles), we can establish the UK-CDI variables' usefulness and quantify its diagnostic ability (sensitivity and specificity) for making predictions in terms of the different criteria of delay at 24 months, see Table 41. Sensitivity levels rise strongly with regards to the increasing cut-off points (i.e. higher percentiles). Specificity is generally very high which is due to the high number of unaffected cases. With an increase in cut-off values, as is the nature of this criterion (to use an increase in percentiles as cut-offs) more children are automatically included in the delay group. This leads to an increase in sensitivity and a decrease in specificity as reported by other authors (e.g., Feldman 2005). Overall, when using the UK-CDI percentile cut-offs, the best results were found for the PLS criterion at 24 months for distinguishing between Production scores at 18 months (80% specificity, 80% sensitivity, good likelihood ratios).

However, instead of using the percentile cut-offs, we can use the ROC-curve output. ROC-curves list different cut-off points (in terms of UK-CDI scores) and their corresponding diagnostic levels. It needs to be decided if any of the two diagnostic criteria (sensitivity, specificity) is more important than the other (an increase in one criterion will automatically decrease the value of the other in most cases) or if both are equally important.

Table 41. *Sensitivity and specificity levels for significant test variables (at 12 and 18 months) per language delay criterion at 24 months*

Criterion	Significant test variables in ROC-curve	Sensitivity	Specificity	LR +	LR -
Rescorla criterion					
	Comprehension 18m				
	< 76 (5th percentile)	27%	100%	very high	0.73
	< 114 (10th percentile)	27%	89%	2.45	0.82
	< 176 (25th percentile)	46%	34%	0.70	1.59
	Production 18m				
	< 3 (5th percentile)	0%	97%	0.00	1.03
	< 7 (10th percentile)	18%	92%	2.25	0.89
	< 17 (25th percentile)	64%	83%	3.76	0.43
	alternative cut-off: 21.5 words	**82%**	**78%**	**3.73**	**0.23**
Vocabulary below 20th percentile AND/OR no word combinations					
	Production 12m				
	< 0 (5th percentile)	29%	85%	1.93	0.84
	< 1 (10th percentile)	59%	80%	2.95	0.51
	< 2 (25th percentile)	59%	69%	1.90	0.59
	Comprehension 18m				
	< 76 (5th percentile)	19%	100%	very high	0.81

	< 114 (10th percentile)	31%	92%	3.88	0.75
	< 176 (25th percentile)	56%	70%	1.87	0.63
	alternative cut-off: 164.5 words	**56%**	**75%**	**2.24**	**0.59**
	alternative cut-off: 205.5 words	**81%**	**65%**	**2.31**	**0.29**
	Production 18m				
	< 3 (5th percentile)	6%	98%	3.00	0.96
	< 7 (10th percentile)	31%	97%	10.33	0.71
	< 17 (25th percentile)	63%	87%	4.85	0.43
	alternative cut-off: 22.5 words	**88%**	**83%**	**5.18**	**0.14**
	Phrases Understood 18m				
	alternative cut-off: 23.5 phrases	**75%**	**53%**	**1.60**	**0.47**
PLS criterion					
	Comprehension 18m				
	< 76 (5th percentile)	20%	97%	6.67	0.82
	< 114 (10th percentile)	60%	90%	6.00	0.44
	< 176 (25th percentile)	80%	68%	2.50	0.29
	alternative cut-off: 181 words	**100%**	**68%**	**3.13**	**0.00**
	Production 18m				
	< 3 (5th percentile)	20%	99%	20.00	0.81
	< 7 (10th percentile)	20%	92%	2.50	0.87
	< 17 (25th percentile)	80%	80%	4.00	0.25
	alternative cut-off: 22.5 words	**100%**	**73%**	**3.70**	**0.00**
	Gestures 18m				
	< 26 (5th percentile)	20%	99%	20.00	0.81
	< 30 (10th percentile)	20%	96%	5.00	0.83
	< 38 (25th percentile)	60%	86%	4.29	0.47
	alternative cut-off: 38.75 gestures	**80%**	**86%**	**5.71**	**0.23**
	Phrases Understood 18m				
	alternative cut-off: 19.5 phrases	**80%**	**73%**	**2.96**	**0.27**

From looking at the *Coordinates of the Curve* table in the ROC-curve output (Appendix 16), we can check if by using a different cut-off score we can improve the diagnostic levels without decreasing the specificity or sensitivity values reached in the previous analyses. For some analyses, there was no improvement possible (e.g. analyses in Rescorla's criterion or Production at 12 months in second criterion).

This shows that for the UK-CDI at around 12 months, the prediction of Production at 12 months cannot be improved by using a different cut-off score. The best possible cut-off for Production at 12 months is the 10th percentile. However, the sensitivity is too low to be useful for clinical purposes. Furthermore, Production at 12 months could not predict language status at 18 or 24 months after using a binary logistic regression in which a step-wise entry was employed to control for the variation in age at 12 months (see Appendix 17 for results). These results show that at around 12 months the UK-CDI Production cannot make reliable predictions about future language status (delay vs. no delay).

There was some improvement possible if only one value increased, and the other remained the same (Comprehension at 18 months in second criterion, Comprehension at 18 months in PLS criterion) or both could be increased (Production at 18 months in second criterion). In some instances, by decreasing the specificity levels slightly the sensitivity could be heightened strongly (Production at 18 months in Rescorla's criterion, Production at 18 months in second criterion, Production at 18 months and Gestures at 18 months in PLS criterion). Table 41 above shows that by choosing a slightly different (usually higher) cut-off score, in most cases at least one or even both levels could be improved.

By increasing the gesture cut-off incrementally, the diagnostic levels became clinically relevant, see Table 41. The sensitivity level increased a lot and the specificity decreased a little which is a reasonable approach according to Westerlund (2006). It should be noted that the Gesture scores include full and half numbers, thus 38 and 38.5 exist. The best cut-off score was 38.75, thus the distinction between 38 and 38.5 seems important for Gesture scores at 18 months. This is because the sensitivity levels increased so rapidly with such a little change in Gesture scores. Further research should check if this result can be replicated.

It can also be seen that the sensitivity levels for Comprehension and Production at 18 months for the PLS criterion could be raised to 100% with specificity levels below 80%. This enables us to correctly identify all children with a language delay rather than just 80% but staying at a relatively low specificity level. This means that whilst we can correctly detect the children with a delay (N=5), the number of incorrectly diagnosed children is high (N=25) in terms of Comprehension scores at 18 months. This number is slightly better for Production at 18 months, if we assume that 32% (N=25) of the lowest ability children are classified as delayed at 18 months (using a cut-off score of 22.5), this results in 20 children being over-identified at 18 months (26%) whilst all children with a later delay were correctly identified (6%). It should be considered that these children who were incorrectly identified had low scores and most of them had low scores six months later but they were above the cut-off for delay.

It has been suggested that at least 80% specificity and sensitivity should be reached to be useful for clinical purposes (Westerlund et al., 2006). As expected, when using the cut-offs for likelihood ratios explained above (Dollaghan, 2013), the decisions about the clinical usefulness generally yield the same results.

The current data shows that cut-off scores had to be higher than the 20th percentile to make clinically meaningful long-term predictions, see Table 41 above. Firstly, it is advised that rather than employing the typically used 10th or 20th percentile to detect language delay concurrently also for making predictions about later language, clinicians should be advised to use the 25th percentile (17 words) or the slightly higher cut-off of 22.5 words (for the production scale at around 18 months). Second, the chosen cut-off points should depend on the follow-up measures generally used in each local authority. This is important as the effectiveness of the UK-CDI categories as predictors and their cut-off points varied strongly depending on the outcome variable (i.e. instrument with criterion of language delay).

Overall the data showed that the UK-CDI used at around 12 months had a limited ability to predict future language delay. Out of all UK-CDI categories (Production, Comprehension, Gestures and Phrases Understood) at 12 months, Production at 12 months had the best prediction but the diagnostic test values were poor which means it should not be used for detecting individual children with a language delay. At 18 months, however, Production and Gestures were clinically useful predictors for language delay at 24 months particularly for the second and PLS criteria. The next analysis investigates if the diagnostic levels remain at the same level at 36 months and if the same children are detected at 36 months in comparison to 24 months.

8.2.3. Determining the criteria for language delay at 36 months

At 36 months, the ASQ communication subscale and the 3-year language instrument were used. According to the ASQ there were 4 children (4.9%) with a delay with regards to the ASQ communication subscale (below 2 SD from mean) and an additional 6 children (7.3%) who should be monitored (between 1 - 2 SDs from the mean) and could be classified as mildly delayed. To create a cut-off point on the vocabulary scale of the 3-year parent report language measure, we chose the 19th percentile (41.8 words) as this contained all children with a mild to severe (below 2 SD) delay on the ASQ. However, it would also *falsely* identify five children with no delay on the ASQ. Furthermore, three children with Production scores around the median, but low grammar skills (children only use simple sentences and no more developed language (no points on the 12-question-scale which includes questions about language comprehension, semantics and syntax)) on the same 3-year parent report booklet would neither be identified using this production cut-off, nor by the ASQ which categorised them as typical.

If we use the 10th percentile (25.5 words) as the criterion for delay, this would include three out of the four children with a severe delay on the ASQ. Again, the same children with slow grammar development would not be detected by this criterion or the ASQ. It should also be mentioned that those children (N=8; 9.8%) below the 10th percentile Production used either very simple sentences (N=4), slightly more complex utterances (N=3) and complex phrases but not yet frequently (N=1). Out of those three children who were delayed on the ASQ and had Production scores below the 10th percentile, only one also used simple sentences and the other two used slightly more complex grammar.

It was decided to include criteria that differed a lot in terms of the number of children it detected as delayed. This was important as different language assessments use different cut-offs for determing delay at 3 years (see more information in the discussion). Hence, the use of the following criteria would give a better overview of the overall direction of the data. The three cut-offs for delay were:

1) Children below the 19th percentile for Production on the 3-year parent report language measure (N = 15, 18.3%), similar cut-offs were used by other researchers (e.g., von Suchodoletz et al., 2009; Zimmerman et al., 2011)
2) ASQ below 2 SD from mean (N = 4, 4.9%): all children with a severe communication delay
3) ASQ below 1 SD (N = 10, 12.1%): all children with a mild or severe communication delay

8.2.4. ROC-curve analyses between UK-CDI scores and delay status at 36 months

For the first criterion, Comprehension at 18 months ($p = .002$) and Production at 18 months ($p < .001$) and Phrases Understood at 18 months ($p = .002$) were significant. None of the test variables were significant on the ASQ when its delay criterion (-2 SD from mean) was used. However, when using the less stringent criterion on the ASQ (-1 SD from mean) tests were significant for Comprehension ($p = .036$), Production ($p = .007$) and Phrases Understood ($p = .011$) at 18 months, meaning that the UK-CDI (at 18 months) can be used for prediction of outcomes on this criterion.

Table 42. *Sensitivity and specificity levels for significant test variables (at 18 months) per language delay criterion at 36 months*

Criterion	Significant test variables in ROC-curve	Sensitivity	Specificity	LR +	LR -
Production below 19th percentile					
	Comprehension 18m				
	< 76 (5th percentile)	7%	99%	7.00	0.94
	< 114 (10th percentile)	36%	92%	4.50	0.70
	< 176 (25th percentile)	64%	76%	2.67	0.47
	Production 18m				
	< 3 (5th percentile)	7%	99%	7.00	0.94
	< 7 (10th percentile)	29%	96%	7.25	0.74
	< 17 (25th percentile)	64%	86%	4.57	0.42

	alternative cut-off: 23.5 words	**86%**	**80%**	**4.30**	**0.18**
	Phrases Understood 18m				
	alternative cut-off: 20.5 phrases	**71%**	**76%**	**2.96**	**0.38**
ASQ communication mild and severe delay					
	Comprehension 18m				
	< 76 (5th percentile)	11%	97%	3.67	0.92
	< 114 (10th percentile)	44%	92%	5.50	0.61
	< 176 (25th percentile)	56%	68%	1.75	0.65
	Production 18m				
	< 3 (5th percentile)	11%	99%	11.00	0.90
	< 7 (10th percentile)	22%	93%	3.14	0.84
	< 17 (25th percentile)	67%	83%	3.94	0.40
	alternative cut-off: 21.5 word	**78%**	**78%**	**3.55**	**0.28**
	Phrases Understood 18m				
	alternative cut-off: 20.5 phrases	**78%**	**73%**	**2.89**	**0.30**

Overall, the UK-CDI at 12 months could not be used to predict language status two years later at 36 months. The results from these analyses also show that in contrast to Comprehension and Phrases Understood, only Production reached levels of diagnostic ability for for clinical usefulness (at or above 80% for specificity and sensitivity) on the 19th percentile production criterion using the alternative cut-off value of 23.5 words on the UK-CDI at 18 months, see Table 42 above. Using this cut-off of 23.5 words on the UK-CDI at 18 months, it categorised 26 children as delayed (31.7%) which was a lot higher than expected for children at around 24 months (Law, Boyle, Harris, Harkness, & Nye, 2000b). Nevertheless, this cut-off could identify most children with a delay depending on the criterion used at 24 months (between 81.8% - 100%)[49] and at 36 months (between 75% - 86.7%)[50]. However, this also meant that out of the 26 children many children were

[49] Rescorla: 9/11= 81.8%, 2nd criterion: 15/17 = 88.2%, PLS: 5/5 = 100%
[50] 19th percentile:13/15 = 86.7%, severe ASQ delay: 3/4 =75%, ASQ mild and severe combined: 8/10 = 80%

not delayed at 24 months[51] (between 42.3%- 80.8%) or 36 months[52] (between 50% – 88.5%) depending on the criterion used, thus these children were over-identified initially.

Last, it was investigated further if the children who were identified as delayed at 18 months and caught up by 36 months (transient delay) differed from children with persistent delays (delay at 18 and 36 months) in terms of participant characteristics. However, we did not find any significant differences between the two groups, see Table 43 below.

Table 43. *Differences between children with a persistent delay (between 18 - 36 months) and children with transient delay (no signs of delay at 36 months)*

	persistent delay (N=13)	transient delay (N=13)	Sig.*/**
male gender	9	10	χ2 (1, n = 26) = .00, *p* = 1.0, *phi* = .09
family risk	3	3	χ2 (1, n = 26) = .00, *p* = 1.0, *phi* = .00
first child	8	8	χ2 (1, n = 26) = .00, *p* = 1.0, *phi* = .00
ear infections	0	2	χ2 (1, n = 26) = .54, *p* = .46, *phi* = -.29
maternal education (A-level or below)	5	3	*t* (24) = -.95, *p* = .35, eta squared = .04
income (at or below £24 000)	3	4	*t* (24) = -.65, *p* = .52, eta squared = .02
cognitive scores 18 months below (at or below 85)	1	1	*t* (24) = -.90, *p* = .38, eta squared = .03
cognitive scores 24 months below (at or below 85)	5	1	*t* (24) = -1.66, *p* = .11, eta squared = .10
motor scores 18 months below (at or below 85)	3	1	*t* (24) = -1.29, *p* = .21, eta squared = .06
motor scores 24 months below (at or below 85)	1	0	*t* (24) = -1.35, *p* = .60, eta squared = .07

*Chi-square test for Independence (With Yates Continuity Correction) for two categorical variables
**Independent t-test for continuous data as dependent variables

8.3. DISCUSSION

Analysis 4 aimed to predict children's language status at 24 and 36 months from scores collected at around 12 and 18 months on an individual rather than group level. The criteria used for

[51] Rescorla: N = 17, 2nd criterion: N = 11, PLS: N= 21
[52] 19th percentile: N = 13, severe ASQ delay: N = 23, ASQ mild and severe combined: N = 18

determining language delay were either set by standardised assessments or created with support from the literature.

UK-CDI scores at around 12 months were unable to make clinically useful predictions at 24 or 36 months. Similar results were found by Sachse et al. (2007) who studied the predictive validity of the German ELFRA-I and found insufficient sensitivity rates of 52% compared to 59% in our data.

In contrast, UK-CDI scores collected at 18 months could significantly predict language outcomes at 24 months and up to 36 months if prevalence rates of language delay were assumed to be high. As found in previous studies, specificity levels were generally very high, but sensitivity was lower. This shows that the ability of typical children is likely to remain in the normal range (specificity) whilst the ability of slowly developing children can be predicted less well as some remain slow and others catch up with their typical peers (sensitivity) (Feldman et al., 2005).

At 36 months compared to 24 months, language is a lot more complex and can be assessed across all linguistic dimensions (morphology, phonology, phonetics, pragmatics, semantics, and syntax). However, the different instruments used at 3 years (ASQ and 3-year-language measure) seemed to measure similar underlying concepts as there was a large overlap. Children who had low scores on both forms at 3 years and low Production scores at 18 months could identify most children with a delay at 3 years, no matter which instrument was used. Again, there was the problem with over-identification of children at 18 months who did not show signs of a delay at 3 years. Some authors use the lowest 10th percentile as the criterion of language delay at 24 months on the CDI (Dale et al., 1998; Ellis Weismer, 2007), but this would not provide clinical useful predictions, whilst others use less stringent criteria around the 20th percentile (e.g., Heilmann, Ellis Weismer, Evans, & Hollar, 2005) which could make significant predictions in our data. However, at 36 months most studies report that less than 10% of children have a language delay (see summary of delay at 36 months literature below) which means that the criterion of delay becomes a lot narrower over time (from around 20% to less than 10%). Thus, due to the way cut-offs are set in practice, a lot of children are identified at first and later do not fall below the cut-off score anymore, even though their scores often remain low, but above the cut-off. However, if we allow a similar percentage of delay over time, we could show that diagnostic criteria would even work up to 36 months with acceptable sensitivity and specificity ratings at or above 80%.

It also needs to be noted that for now, we need to be careful with the definition of language delay at 3 years. This is because the ASQ has so far only been normed and standardised for US American children. While the 3-year language measure has also been created for British children, so far no norms exist at this age. Previous research reported prevalence rates for language delay at 3 years between 2.3 and 7.6% (Law, Boyle, Harris, Harkness, & Nye, 1998; Law, Boyle, Harris, Harkness, & Nye, 2000b). More recent screening tests chose somewhat higher cut-offs to include at-risk children. For example, Bleses et al. (2010) developed a new language screening instrument for 3-year-olds (Danish SI-3) and created gender specific cut-off scores for children with low ability (5^{th} percentile) and at-risk children (between 6^{th}-13^{th} percentiles). In addition, the SBE-3-KT instrument which assesses vocabulary and grammar ability in 3-year-old German children uses the 16^{th} percentile as the cut-off (von Suchodoletz et al., 2009). The Preschool Language Scale 5-UK uses a standard score of 85 as the cut-off across all age groups which is the equivalent of the 16^{th} percentile (Zimmerman et al., 2011). This shows that it is difficult to give an exact estimation of language delay at 3 years as the prevalence rates depend on the instruments' use of cut-off points which then differ in the amount of over- and under-identification test developers were allowing for (Feldman et al., 2005; von Suchodoletz et al., 2009). For the current study, we used the 19^{th} percentile as a cut-off on the 3-year language measure. This included all children with a mild and severe delay on the ASQ. From the prevalence rates reported here, it seems that this cut-off (19^{th} percentile) was somewhat high; however, when we used lower cut-offs the predictions did not reach clinically useful levels at this age.

Overall, when using cut-off points *recommended by the literature*, the UK-CDI was unable to give accurate predictions for clinical use up to 24 or 36 months which is supported by other research (Dollaghan, 2013; Feldman et al., 2005; Unhjem et al., 2015; Westerlund et al., 2006). Nevertheless, the data shows that long-term predictions from 18 to 24 or 36 months were possible if the UK-CDI score cut-off was higher than the 20th percentile (a frequently used cut-off for language delay) and if the prevalence rate of delay at 36 months was set higher than assumed at this age. However, even when clinically useful cut-off values were identified with diagnostic criteria in the acceptable range, some children would still be over-identified at 18 months which means they would show signs of a delay at 18 months but would not fall below the cut-off for delay later during development. Similar results were found by Thal et al. (2013). When investigating potential differences between children with persistent delays and those with transient delays (over-identified at 18 months) in terms of participant characteristics, we did not find significant differences between the two groups in the current sample. Other research also found that characteristics such as ear infections or maternal education could not significantly improve the prediction for those children with persistent delays in comparison to those with transient delays (Dale et al., 2003).

It may be argued that it is a political and monetary question, whether a society can afford to support many more children than just those who require support in the future. If universal screening was implemented at 18 months using the UK-CDI with more restrictive criteria for delay, there is a possibility to disregard a substantial group of children who will go on to develop a language deficit in the future.

In contrast, if less restrictive criteria were used more children would be identified and also over-identified, but our data suggests that many of these children still remain less skilled than those children with average and above average skills (see results in Analysis 1 and 2). Therefore, it is likely that these children who are over-identfied would also benefit from early intervention.

Early intervention often consists of parent-implemented language intervention (e.g., Buschmann et al., 2009; Buschmann et al., 2015; Kaiser & Roberts, 2011; M. Y. Roberts & Kaiser, 2011; M. Y. Roberts & Kaiser, 2015) which typically involves in-person training that may be costly. To keep costs low and offer a service to all parents, it may also be possible to develop NHS approved apps or podcasts (NHS, 2018). Whilst these products could be available to everyone, certain features could be made accessible only to those parents of children with an identified delay (e.g. monitoring of progress and communication with the health visitor via the app). Other researchers have also suggested that low cost and parent-based interventions should be implemented already during the toddler period (Bleses et al., 2016). This is because even though most children catch up with their peers at kindergarten age demonstrating scores within the normal range on standardised language tests (Thal et al., 2013), late-talkers are at continued risk for low educational attainment (Bleses et al., 2016).

9. Summary and Conclusion

The aim of this PhD was to examine the predictive validity of the newly developed UK-CDI (Alcock et al., 2017). We investigated in Analysis 1 the associations between UK-CDI scores at 12 and 18 months with language scores up to 3 years and found similar results as reported in previous CDI research. The strongest long-term correlations were found between UK-CDI scores (except for Gestures) at 18 months and later language scores.

In Analysis 2, more in depth analyses revealed that the language of those children with high skills (particularly for Production ability) at 18 months remained most stable over time which means they had significantly better language outcomes compared to those children who were slower at developing language initially. In contrast, children with low ability showed more variability in outcome.

Furthermore, we found that the UK-CDI at 12 months could not explain any more variance in later language above and beyond the UK-CDI at 18 months, whilst some biological and environmental influences could explain some significant additional variance after controlling for UK-CDI scores. However, these influences were not consistent in explaining future language, in Analysis 3.

As early language was the best predictor for later language, we used ROC-curve analyses in Analysis 4 to predict language delay in individual children. As found in previous CDI research, prediction was not possible when using commonly employed cut-off points for delay. Instead, it is advised to use higher cut-off points (> 20th percentile) at 18 months, as then acceptable sensitivity and specificity levels were found up to 3 years. At 3 years, Production scores at 18 months and the criterion for Production delay at 3 years (using the 3-year parent report measure) gave the highest classification accuracy. Again, the strongest continuity was also found for the same category over time.

In sum, it has been demonstrated that the UK-CDI has predictive validity on a group level as we found continuity of language ability across time. When using common cut-off points employed in the literature, the UK-CDI cannot be advised for the early identification of language delay for individual children. However, if higher cut-off points are accepted which automatically over-identify many children, clinically meaningful predictions can be made for those children who remain delayed at least up to 3 years old when first identified at 18 months using the UK-CDI.

Future research could establish the predictive validity of the UK-CDI separately for children at familial risk for dyslexia or speech and language problems. Previous research has suggested that children at risk for dyslexia show more stability in their language development (Unhjem et al., 2015), thus it is possible that the UK-CDI when used during the early stages of language development could be successful at correctly identifying individual children who will go on to develop language related problems in the furture.

Furthermore, it would be useful to recruit a large number of children with low Production scores at 12 months (below 25th percentile on the UK-CDI) and follow them longitudinally up to at least school age. This would be worthwhile as we found that these low ability children had significantly lower grammar skills at 36 months compared to all other groups. Research suggests that children with initial slow development catch up with their peers on most language related measures at kindergarten age. However, even though they may fall within in the typical range on standardised tests, they may still show significantly lower skills than typically developing children, particularly in newly developing skills such as reading related skills (Thal et al., 2013).

References

Acredolo, L., & Goodwyn, S. (1988). Symbolic gesturing in normal infants. *Child Development, 59*(2), 450-466. doi:10.2307/1130324

Alcock, K. J., Meints, K., & Rowland, C. (2017). *UK-CDI: Words and gestures - preliminary norms and manual.* Unpublished manuscript.

Alcock, K. J., Meints, K., Rowland, C. F., Christopher, A., Just, J., & Brelsford, V. (in prep). The UK communicative development inventory: Words and gestures.

Alcock, K. J., Rimba, K., Holding, P., Kitsao-Wekulo, P., Abubakar, A., & Newton, C. R. J. C. (2015). Developmental inventories using illiterate parents as informants: Communicative development inventory (CDI) adaptation for two Kenyan languages. *Journal of Child Language, 42*(4), 763-785. doi:10.1017/S0305000914000403

Alexander, G. M., Wilcox, T., & Woods, R. (2009). Sex differences in infants' visual interest in toys. *Archives of Sexual Behavior, 38*, 427-433.

Anastasi, A., & Urbino, S. (1997). *Psychological testing*. Upper Saddle River, NJ: Prentice Hall.

Arriaga, R., Fenson, L., Cronan, T., & Pethick, S. (1998). Scores on the MacArthur communicative development inventory of children from low- and middle-income families. *Applied Psycholinguistics, 19*, 209-223.

Bailey, D. B., & Wolery, M. (1989). *Assessing infants and preschoolers with handicaps*. Columbus, OH: Merrill.

Bates, E., Beeghly, M., Bretherton, I., Harris, C., Marchman, V. A., McNew, S., . . . Whitesell, K. (1986). *The language and gesture inventory.* Unpublished checklist, University of California, San Diego.

Bates, E., Bretherton, I., & Snyder, L. (1988). *From first words to grammar: Individual differences and dissociable mechanisms*. Cambridge, MA: Cambridge University Press.

Baumwell, L., & Tamis, C. S. (1997). Maternal verbal sensitivity and child language comprehension. *Infant Behavior and Development, 20*(2), 247-258.

Bavin, E. L., & Bretherton, L. (2013). The early language in Victoria study: Late talkers, predictors, and outcomes. In L. A. Rescorla, & P. S. Dale (Eds.), *Late talkers language development, interventions and outcomes* (pp. 3-21). London: Paul H. Brookes Publishing.

Bavin, E. L., Prior, M., Reilly, S., Bretherton, L., Williams, J., Eadie, P., . . . Ukoumunne, O. C. (2008). The early language in Victoria study: Predicting vocabulary at age one and two years from gesture and object use. *Journal of Child Language, 35*(3), 687-702.

Baxendale, J., & Hesketh, A. (2003). Comparison of the effectiveness of the Hanen parent programme and traditional clinic therapy. *International Journal of Language & Communication Disorders, 38*(4), 397 - 415.

Bayley, N. (2006a). *Bayley scales of infant and toddler development – third edition: Administration manual*. San Antonio, TX: Harcourt Assessment.

Bayley, N. (2006b). *Bayley scales of infant and toddler development – third edition: Technical manual*. San Antonio, TX: Harcourt Assessment.

Bayley, N. (2010). *Bayley scales of infant and toddler development third edition Nancy Bayley UK and ireland supplement manual*. London: Pearson Assessment.

Bedford, H., Walton, S. & Ahn, J. (2013). Measures of child development: A review. Retrieved from https://www.ucl.ac.uk/cpru/documents/review_of_measures_of_child_development

Beeghley, M. (1998). Emergence of symbolic play: Perspectives from typical and atypical development. In J. A. Burack, R. M. Hodapp & E. F. Zigler (Eds.), *Handbook of mental retardation and development* (pp. 240-288). Cambridge: University Press.

Bergelson, E., & Swingley, D. (2012). At 6-9 months, human infants know the meanings of many common nouns. *Proceedings of the National Academy of Sciences of the United States of America, 109*(9), 3253.

Berglund, E., & Eriksson, M. (2000). Communicative development in Swedish children 16-28 months old: The Swedish early communicative development inventory-words and sentences. *Scandinavian Journal of Psychology, 42*(2), 133-144.

Berglund, E., Eriksson, M., & Westerlund, M. (2005). Communicative skills in relation to gender, birth order, childcare and socioeconomic status in 18-month-old children. *Scandinavian Journal of Psychology, 46*(6), 485-491. doi:10.1111/j.1467-9450.2005.00480.x

Bishop, D. V. M., & Edmundson, A. (1987). Specific language impairment as a maturational lag: Evidence from longitudinal data on language and motor development. *Developmental Medicine and Child Neurology, 29*, 442-459.

Bishop, D. V. M., Price, T. S., Dale, P. S., & Plomin, R. (2003). Outcomes of early language delay: II. etiology of transient and persistent language difficulties. *Journal of Speech, Language & Hearing Research, 46*(3), 561-575.

Blencowe, H., Cousens, S., Oestergaard, M. Z., Chou, D., Moller, A., Narwal, R., . . . Lawn, J. E. (2012). National, regional, and worldwide estimates of preterm birth rates in the year 2010 with time trends since 1990 for selected countries: A systematic analysis and implications. *The Lancet, 379*(9832), 2162-2172.

Bleses, D., Basboll, H., Lum, J., & Vach, W. (2011). Phonology and lexicon in a cross-linguistic perspective: The importance of phonetics--A commentary on Stoel-Gammon's "relationships between lexical and phonological development in young children". *Journal of Child Language, 38*(1), 61-68.

Bleses, D., Makransky, G., Dale, P. S., HØjen, A., & Ari, B. (2016). Early productive vocabulary predicts academic achievement 10 years later. *Applied Psycholinguistics, 37*(6), 1461-1476. doi:10.1017/S0142716416000060

Bleses, D., Vach, W., Jørgensen, R. N., & Worm, T. (2010). The internal validity and acceptability of the Danish SI-3: A language-screening instrument for 3-year-olds. *Journal of Speech, Language & Hearing Research, 53*(2), 490-507. doi:1092-4388(2009/08-0132

Bleses, D., Vach, W., Slott, M., Wehberg, S., Thomsen, P., Madsen, T. O., & Basboll, H. (2008). The Danish communicative developmental inventories: Validity and main developmental trends. *Journal of Child Language, 35*(3), 651-670.

Bleses, D., Vach, W., Slott, M., Wehberg, S., Thomsen, P., Madsen, T. O., & Basboll, H. (2008). Early vocabulary development in Danish and other languages: A CDI-based comparison. *Journal of Child Language, 35*(3), 619-650.

Bochner, S., & Jones, J. (2003). *Child language development learning to talk* (2nd ed.). London: Whurr Publishers Ltd. Retrieved from http://proxy.library.lincoln.ac.uk/login?url=http://search.ebscohost.com/login.aspx?direct=true&db=edb&AN=74878542&site=eds-live&scope=site

Boets, B., Vandermosten, M., Poelmans, H., Luts, H., Wouters, J., & Ghesquiere, P. (2011). Preschool impairments in auditory processing and speech perception uniquely predict future reading problems. *Research in Developmental Disabilities, 32*(2), 560-570. doi:10.1016/j.ridd.2010.12.020

Bopp, K. D., & Pat, M. (2011). Prelinguistic predictors of language development in children with Autism Spectrum Disorders over four-five years. *Journal of Child Language, 38*(3), 485-504.

Bornstein, M. H., & Haynes, O. M. (1998). Vocabulary competence in early childhood: Measurement, latent construct, and predictive validity. *Child Development, 69*(3), 654-671. doi:10.1111/j.1467-8624.1998.tb06235.x

Bornstein, M. H., Haynes, O. M., & Painter, K. (1998). Sources of child vocabulary competence: A multivariate model. *Journal of Child Language, 25*(2), 367-393.

Bornstein, M. H., Leach, D. B., & Haynes, O. M. (2004). Vocabulary competence in first- and secondborn siblings of the same chronological age. *Journal of Child Language, 31*(4), 855-873.

Bornstein, M. H., Putnick, D. L., & De Houwer, A. (2006). Child vocabulary across the second year: Stability and continuity for reporter comparisons and cumulative score. *First Language, 26*(3), 299-316.

Bradley, R.H., Corwyn, R.F., McAdoo, H.P., & Garcia Coll, C. (2001). The home environments of children in the United States. Part I: Variations by age, ethnicity, and poverty status. *Child Development, 72*(6), 1844–67.

Brownell, R. (2000a). *Expressive one-word picture vocabulary test (EOWPVT)*. Novato, CA: Academic Therapy Publications.

Brownell, R. (2000b). *Receptive one-word picture vocabulary test (ROWPVT)*. Novato, CA: Academic Therapy Publications.

Buschmann, A., Jooss, B., Rupp, A., Dockter, S., Blaschtikowitz, H., Heggen, I., & Pietz, J. (2008). Children with developmental language delay at 24 months of age: Results of a diagnostic work-up. *Developmental Medicine & Child Neurology, 50*(3), 223-229. doi:10.1111/j.1469-8749.2008.02034.x

Buschmann, A., Jooss, B., Rupp, A., Feldhusen, F., Pietz, J., & Philippi, H. (2009). Parent based language intervention for 2-year-old children with specific expressive language delay: A randomised controlled trial. *Archives of Disease in Childhood, 94*(2), 110-116.

Buschmann, A., Multhauf, B., Hasselhorn, M., & Pietz, J. (2015). Long-term effects of a parent-based language intervention on language outcomes and working memory for late-talking toddlers. *Journal of Early Intervention, 37*(3), 175-189 15p. doi:10.1177/1053815115609384

Caldera, Y. M., Huston, A. C., & O'Brien, M. (1989). Social interactions and play patterns of parents and toddlers with feminine, masculine, and neutral toys. *Child Development, 60*(1), 70-77.

Campbell, A., Shirley, L., Heywood, C., & Crook, C. (2000). Infants' visual preference for sex-congruent babies, children, toys and activities: A longitudinal study. *British Journal of Developmental Psychology, 18*, 479-498.

Can, D. D., Ginsburg-Block, M., Golinkoff, R. M., & Hirsh-Pasek, K. (2013). A long-term predictive validity study: Can the CDI short form be used to predict language and early literacy skills four years later? *Journal of Child Language, 40*(04), 821-835. doi:10.1017/S030500091200030X

Caselli, M. C. (1995). A cross-linguistic study of early lexical development. *Cognitive Development, 10*(2), 159-199.

Castellanos, I., Pisoni, D. B., Kronenberger, W. G., & Beer, J. (2016). Early expressive language skills predict long-term neurocognitive outcomes in cochlear implant users: Evidence from the MacArthur-Bates communicative development inventories. *American Journal of Speech-Language Pathology, 25*(3), 381-392. doi:10.1044/2016_AJSLP-15-0023

Cattani, A., Abbot-Smith, K., Farag, R., Krott, A., Arreckx, F., Dennis, I., & Floccia, C. (2014). How much exposure to English is necessary for a bilingual toddler to perform like a monolingual peer in language tests? *International Journal of Language & Communication Disorders, 49*(6), 649-671. doi:10.1111/1460-6984.12082

Chan, A., Meints, K., Lieven, E., & Tomasello, M. (2010). Young children's comprehension of English SVO word order revisited: Testing the same children in act-out and intermodal preferential looking tasks. *Cognitive Development, 25*, 30-45. doi:10.1016/j.cogdev.2009.10.002

Chapman, R. S. (1978). Comprehension strategies in children. In J. F. Kavanagh, & W. Strange (Eds.), *Speech and language in the laboratory, school and clinic* (pp. 308-327). Cambridge, MA: MIT Press.

Ciccone, N., Hennessey, N., & Stokes, S. F. (2012). Community-based early intervention for language delay: A preliminary investigation. *International Journal of Language & Communication Disorders, 47*(4), 467-470 4p. doi:10.1111/j.1460-6984.2012.00149.x

Cochet, H., & Byrne, R. W. (2016). Communication in the second and third year of life: Relationships between nonverbal social skills and language. *Infant Behavior and Development, 44*, 189-198. doi:10.1016/j.infbeh.2016.07.003

Collisson, B. A., Graham, S. A., Preston, J. L., Rose, M. S., McDonald, S., & Tough, S. (2016). Risk and protective factors for late talking: An epidemiologic investigation. *Journal of Pediatrics, 172*, 168-174. doi:10.1016/j.jpeds.2016.02.020

Colonnesi, C., Stams, G. J. J. M., Koster, I., & Noom, M. J. (2010). The relation between pointing and language development: A meta-analysis. *Developmental Review, 30*(4), 352-366.

Corkum, V., & Dunham, P. (1996). The communicative development inventory-WORDS short form as an index of language production. *Journal of Child Language, 23*(03), 515-528. doi:10.1017/S0305000900008928

Crnic, K. A., Ragozin, A. S., Greenberg, M. T., Robinson, N. M., & Basham, R. B. (1983). Social interaction and developmental competence of preterm and full-term infants during the first year of life. *Child Development,* (5), 1199-1210.

Dale, P. S., & Hayiou-Thomas, M. E. (2013). Outcomes for late talkers: A twin study. In L. A. Rescorla, & P. S. Dale (Eds.), *Late talkers language development, interventions, and outcomes* (pp. 241-257). London: Paul H. Brookes Publishing.

Dale, P. S., McMillan, A. J., Hayiou-Thomas, M. E., & Plomin, R. (2014). Illusory recovery: Are recovered children with early language delay at continuing elevated risk? *American Journal of Speech-Language Pathology, 23*(4), 437-447. doi:10.1044/2014

Dale, P. S., Price, T. S., Bishop, D. V. M., & Plomin, R. (2003). Outcomes of early language delay: I. predicting persistent and transient language difficulties at 3 and 4 years. *Journal of Speech, Language & Hearing Research, 46*(3), 544-560.

Dale, P. S., Reznick, J. S., Thal, D. J., & Marchman, V. A. (2001). *A parent report measure of language development for three-year-olds.* Unpublished manuscript, University of Missouri, Columbia.

Dale, P. S., Simonoff, E., Bishop, D. V. M., & Plomin, R. (1998). Genetic influence on language delay in two-year-old children. *Nature Neuroscience, 1*(4), 324-328.

de Carvalho, C. A. F., Kida, A. D. B., Capellini, S. A., & de Avila, C. R. B. (2014). Phonological working memory and reading in students with dyslexia. *Frontiers in Psychology, 5*(746) doi:10.3389/fpsyg.2014.00746

Deckers, S. R. J. M., Van Zaalen, Y., Mens, E. J. M., Van Balkom, H., & Verhoeven, L. (2016). The concurrent and predictive validity of the Dutch version of the communicative development inventory in children with down syndrome for the assessment of expressive vocabulary in verbal and signed modalities. *Research in Developmental Disabilities, 56*, 99-107. doi:http://dx.doi.org.proxy.library.lincoln.ac.uk/10.1016/j.ridd.2016.05.017

Desmarais, C., Sylvestre, A., Meyer, F., Bairati, I., & Rouleau, N. (2008). Systematic review of the literature on characteristics of late-talking toddlers. *International Journal of Language & Communication Disorders, 43*(4), 361-390.

Diaz, M. F. L., Fonseca, A. M. G., Bohorquez, D. M. G., Guecha, C. M., & Sellabona, E. S. (2011). Normativization to the Spanish-Colombia to the CDI- MacArthur-Bates communicative development inventories. *Revista Latinoamericana De Psicologia, 43*(2), 241-254.

Dionne, G., Dale, P. S., Boivin, M., & Plomin, R. (2003). Genetic evidence for bidirectional effects of early lexical and grammatical development. *Child Development, 74*(2), 394-412. doi:10.1111/1467-8624.7402005

D'Odorico, L., Majorano, M., Fasolo, M., Salerni, N., & Suttora, C. (2011). Characteristics of phonological development as a risk factor for language development in Italian-speaking pre-term children: A longitudinal study. *Clinical Linguistics & Phonetics, 25*(1), 53-65. doi:10.3109/02699206.2010.511759

Dollaghan, C. (2013). Late talker as a clinical category: A critical evaluation. In L. A. Rescorla, & P. S. Dale (Eds.), *Late talkers language development, interventions, and outcomes* (pp. 91-112). London: Paul H. Brookes Publishing.

D'Souza, D., D'Souza, H., & Karmiloff-Smith, A. (2017). Precursors to language development in typically and atypically developing infants and toddlers: The importance of embracing complexity. *Journal of Child Language, 44*(3), 591-628.

Duff, F. J., Nation, K., Plunkett, K., & Bishop, D. V. M. (2015). Early prediction of language and literacy problems: Is 18 months too early? *Peerj, 3*, e1098-e1098. doi:10.7717/peerj.1098

Duff, F. J., Reen, G., Plunkett, K., & Nation, K. (2015). Do infant vocabulary skills predict school-age language and literacy outcomes? *Journal of Child Psychology and Psychiatry, 56*(8), 848-856.

Edwards, S., Fletcher, P., Gurman, M., Hughes, A., & Letts, C. (1997). *Reynell developmental language scales III (RDLS III) (University of Reading edition)*. Windsor: NFER-Nelson Health & Social Care.

Edwards, S., Letts, C., & Sinka, I. (2011). *The new Reynell developmental language scales*. London: GL-Assessment.

Ellis Weismer, S. (2007). Typical talkers, late talkers, and children with specific language impairment: A language endowment spectrum? In R. Paul (Ed.), *Language disorders from a developmental perspective: Essays in honor of Robin S. Chapman* (pp. 83-102). Mahwah, NJ: Erlbaum.

Elman, J. L., Bates, E., Johnson, M. H., Karmiloff-Smith, A., Parisi, D., & Plunkett, K. (1996). *Rethinking innateness a connectionist perspective on development*. London: MIT Press.

Eriksson, M. (2001). Narratives validate communicative development inventories. *Applied Psycholinguistics, 22*(01), 45-60.

Eriksson, M., & Berglund, E. (1999). Swedish early communicative development inventories: Words and gestures. *First Language, 19*(55), 55-91.

Eriksson, M., Marschik, P. B., Tulviste, T., Almgren, M., Perez Pereira, M., Wehberg, S., . . . Gallego, C. (2012). Differences between girls and boys in emerging language skills: Evidence from 10 language communities. *British Journal of Developmental Psychology, 30*(2), 326-344.

Feldman, H. M., Dale, P. S., Campbell, T. F., Colborn, D. K., Kurs-Lasky, M., Rockette, H. E., & Paradise, J. L. (2005). Concurrent and predictive validity of parent reports of child language at ages 2 and 3 years. *Child Development,* (4), 856-868. doi:10.1111/j.1467-8624.2005.00882.x

Feldman, H. M., Dollaghan, C. A., Campbell, T. F., Kurs-Lasky, M., Janosky, J. E., & Paradise, J. L. (2000). Measurement properties of the MacArthur communicative development inventories at ages one and two years. *Child Development, 71*(2), 310-323.

Fenson, L., Bates, E., Dale, P., Goodman, J., Reznick, J. S., & Thal, D. J. (2000). Measuring variability in early child language: Don't shoot the messenger: [reply]. *Child Development,* (2), 323-328.

Fenson, L., Dale, P. S., Bates, E., Reznick, J. S., Thal, D. J., Jackson-Maldonado, D., . . . Friend, M. (2015). Adaptations in other languages. Retrieved from http://mb-cdi.stanford.edu/adaptations.html

Fenson, L., Dale, P. S., Reznick, J. S., Bates, E., Thal, D. J., & Pethick, S. J. (1994). In Bronson W. C. (Ed.), *Variability in early communicative development*. Chicago, Illinois: The University of Chicago Press.

Fenson, L., Dale, P. S., Reznick, J. S., Thal, D. J., Bates, E., Hartung, J. P., . . . Reilly, J. S. (1993). *The MacArthur communicative development inventories: User's guide and technical manual*. San Diego, CA: Singular Publishing Group.

Fenson, L., Marchman, V. A., Thal, D. J., Dale, P. S., Reznick, J. S., & Bates, E. (Eds.). (2007). *MacArthur-Bates communicative development inventories: User's guide and technical manual* (2nd ed.). Baltimore, MD: Brookes.

Fernald, A., Marchman, V., & Weisleder, A. (2013). SES differences in language processing skill and vocabulary are evident at 18 months. *Developmental Science, 16*(2), 234 – 248.

Field, A. (2017). *Discovering statistics using IBM SPSS* (5th ed.). Los Angeles: SAGE.

Fish, M., & Pinkerman, B. (2003). Language skills in low-SES rural Appalachian children: Normative development and individual differences, infancy to preschool. *Journal of Applied Developmental Psychology, 23*, 539-565. doi:10.1016/S0193-3973(02)00141-7

Floccia, C., Sambrook, T. D., & Delle Luche, C. (2017). Vocabulary of 2-year-olds learning English and an additional language: Norms and effects of linguistic distance. *Monographs of the Society for Research in Child Development,*

Folio, R. M., & Fewell, R. R. (2000). *Peabody developmental motor scales, second edition*. Austin, TX: Pro-Ed.

Forum for Research into Language and Literacy. (2012). *Diagnostic test of word reading processes*. London: GL Assessment.

Friend, M., Schmitt, S. A., & Simpson, A. M. (2012). Evaluating the predictive validity of the computerized comprehension task: Comprehension predicts production. *Developmental Psychology, 48*(1), 136-148. doi:10.1037/a0025511

Fricke, S., Bowyer-Crane, C., Haley, A., Hulme, C., & Snowling, M. (2013). Efficacy of language intervention in the early years. *Journal of Child Psychology and Psychiatry*, *54*(3), 280–290.

Fricke, S., Burgoyne, K., Bowyer-Crane, C., Kyriacou, M., Zosimidou, A., Maxwell, L., … Hulme, C. (2017). The Efficacy of Early Language Intervention in Mainstream School Settings: A Randomized Controlled Trial. *Journal of Child Psychology and Psychiatry*, *58*(10), 1141–1151.

Froiland, J. M., Powell, D. R., Diamond, K. E., & Son, S.H. C. (2013). Neighborhood Socioeconomic Well-Being, Home Literacy, and Early Literacy Skills of At-Risk Preschoolers. *Psychology in the Schools*, *50*(8), 755–769.

Frota, S., Butler, J., Correia, S., Severino, C., Vicente, S., & Vigário, M. (2016). Infant communicative development assessed with the European Portuguese MacArthur–Bates communicative development inventories short forms. *First Language, 36*(5), 525-545.

Gallagher, A., Frith, U., & Snowling, M. J. (2000). Precursors of literacy delay among children at genetic risk of dyslexia. *Journal of Child Psychiatry, 41*(2), 203-213.

Garcia, I., Arratibel, N., Barrena, A., & Ezeizabarrena, M. J. (2008). Adaptation of the MacArthur-Bates inventories to the Basque language: Communicative development in children between 8 and 30 months of age. *Infancia y Aprendizaje, 31*(4), 411-424.

Gathercole, S. E., & Baddeley, A. D. (1996). *The children's test of nonword repetition*. London: Psychological Corporation.

Gemeinsamer Bundesausschuss. (2016). Beschluss des gemeinsamen Bundesausschusses über eine Neufassung der Richtlinien über die Früherkennung von Krankheiten bei Kindern bis zur Vollendung des 6. Lebensjahres (Kinder-Richtlinien): Formale und inhaltliche Überarbeitung (Neustrukturierung). Retrieved from https://www.g-ba.de/downloads/39-261-2287/2015-06-18_2015-08-20_2016-05-19_2016-07-07_Kinder-RL_Neustrukturierung_Neufassung_konsolidiert_BAnz.pdf

Gershkoff-Stowe, L., & Smith, L. B. (1997). A curvilinear trend in naming errors as a function of early vocabulary growth. *Cognitive Psychology, 34*(1), 37-72.

Ghassabian, A., Rescorla, L. A., Henrichs, J., Jaddoe, V. W., Verhulst, F. C., & Tiemeier, H. (2014). Early lexical development and risk of verbal and nonverbal cognitive delay at school age. *Acta Paediatrica, 103*(1), 70-81.

Gilger, J. W., & Kaplan, B. J. (2001). Atypical brain development: A conceptual framework for understanding developmental learning disabilities. *Developmental Neuropsychology, 20*(2), 465-481. doi:10.1207/S15326942DN2002_2

Glennen, S. L. (2007). Predicting language outcomes for internationally adopted children. *Journal of Speech, Language & Hearing Research, 50*(2), 529-549.

Goldin-Meadow, S. (2007). Gesture with speech and without it. In S. C. Duncan, J. Cassell & E. T. Levy (Eds.), *Gesture and the dynamic dimension of language: Essays in honor of david McNeill* (pp. 31-39). Amsterdam: Benjamins.

Greenfield, P. M., Keller, K., Fuligni, A., & Maynard, A. (2003). Cultural pathways through universal development. *Annual Review of Psychology, 54*, 461-490.

Grimm, H., & Doil, H. (2006). *ELFRA Elternfragebögen für die Früherkennung von Risikokindern* (2nd ed.). Göttingen: Hogrefe Verlag.

Guarini, A., Sansavini, A., Fabbri, C., Savini, S., Alessandroni, R., Faldella, G., & Karmiloff-Smith, A. (2010). Long-term effects of preterm birth on language and literacy at eight years. *Journal of Child Language, 37*(4), 865-885.

Guiberson, M. (2008). Validity of a parent vocabulary checklist for young Spanish speaking children of Mexican immigrants. *International Journal of Speech-Language Pathology, 10*(5), 279-286.

Hamadani, J. D., Baker-Henningham, H., Tofail, F., Mehrin, F., Huda, S. N., & Grantham-McGregor, S. M. (2010). Validity and reliability of mothers' reports of language development in 1-year-old children in a large-scale survey in Bangladesh. *Food & Nutrition Bulletin, 31*(2), 198-206.

Hamilton, A., Plunkett, K., & Schafer, G. (2000). Infant vocabulary development assessed with a British communicative development inventory. *Journal of Child Language, 27*(3), 689-705.

Harris, F., Law, J., & Roy, P. (2005). The third implementation of the sure start language measure . *Sure Start Evidence & Research,*

Hart, B., & Risely, T. R. (1995). *Meaningful differences in the everyday experience of young American children*. Baltimore: Paul H. Brookes.

Hayiou-Thomas, M. E., Dale, P. S., & Plomin, R. (2012). The etiology of variation in language skills changes with development: A longitudinal twin study of language from 2 to 12 years. *Developmental Science, 15*(2), 233-249.

Hedrick, D. L., Prather, E. M., & Tobin, A. R. (1984a). *Sequenced inventory of communication development*. Seattle: University of Washington Press.

Hedrick, D. L., Prather, E. M., & Tobin, A. R. (1984b). *Sequenced inventory of communication development — revised*. Seattle, WA: University of Washington Press.

Heilmann, J., Ellis Weismer, S., Evans, J., & Hollar, C. (2005). Utility of the MacArthur–Bates communicative development inventory in identifying language abilities of late-talking and typically developing toddlers. *American Journal of Speech-Language Pathology, 14*, 40-51.

Henrichs, J., Rescorla, L. A., Schenk, J., Schmidt, H. G., Jaddoe, V. W. V., Hofman, A., . . . Tiemeier, H. (2011). Examining continuity of early expressive vocabulary development: The generation R study. *Journal of Speech, Language & Hearing Research, 54*(3), 854-869. doi:10.1044/1092-4388(2010/09-0255)

Hoff, E. (2003). The specificity of environmental influence: Socioeconomic status affects early vocabulary development via maternal speech. *Child Development,* (5), 1368-1378.

Hoff-Ginsberg, E. (1998). The relation of birth order and socioeconomic status to children's language experience and language development. *Applied Psycholinguistics, 19*(04), 603-629. doi:10.1017/S0142716400010389

Hofman, J. M., Sharma, A., & Watts, D. J. (2017). Prediction and explanation in social systems. *Science, 355*(6324), 486-488.

Hohm, E., Jennen-Steinmetz, C., Schmidt, M., & Laucht, M. (2007). Language development at ten months. *European Child & Adolescent Psychiatry, 16*(3), 149-156. doi:10.1007/s00787-006-0567-y

Hohnen, B., & Stevenson, J. (1999). The structure of genetic influences on general cognitive, language, phonological, and reading abilities. *Developmental Psychology, 35*(2), 590-603. doi:10.1037/0012-1649.35.2.590

Horwitz, S. M., Irwin, J. R., Briggs-Gowan, M., Heenan, J. M. B., Mendoza, J., & Carter, A. S. (2003). Language delay in a community cohort of young children. *Journal of the American Academy of Child & Adolescent Psychiatry, 42*(8), 932.

Hsu, H., & Iyer, S. N. (2016). Early gesture, early vocabulary, and risk of language impairment in preschoolers. *Research in Developmental Disabilities, 57*, 201-210. doi:http://dx.doi.org/10.1016/j.ridd.2016.06.012

Igualada, A., Bosch, L., & Prieto, P. (2015). Language development at 18 months is related to multimodal communicative strategies at 12 months. *Infant Behavior and Development, 39*, 42-52.

Iverson, J. M. (2010). Developing language in a developing body: The relationship between motor development and language development. *Journal of Child Language, 37*(2), 229-262.

Iverson, J. M., & Goldin-Meadow, S. (2005). Gesture paves the way for language development. *Psychological Science, 16*(5), 367-371.

Kaduk, K., Bakker, M., Juvrud, J., Gredebäck, G., Westermann, G., Lunn, J., & Reid, V. M. (2016). Semantic processing of actions at 9 months is linked to language proficiency at 9 and 18 months. *Journal of Experimental Child Psychology, 151*, 96-108. doi:http://dx.doi.org.proxy.library.lincoln.ac.uk/10.1016/j.jecp.2016.02.003

Kahan, D., McKenzie, T. L., Smith, L., van Jaarsveld, C. H. M., Llewellyn, C. H., Fildes, A., . . . Fisher, A. (2017). Genetic and environmental influences on developmental milestones and movement: Results from the gemini cohort study. *Research Quarterly for Exercise & Sport, 88*(4), 524-529. doi:10.1080/02701367.2017.1373268

Kaiser, A. P., & Roberts, M. Y. (2011). Advances in early communication and language intervention. *Journal of Early Intervention, 33*(4), 298-309. doi:10.1177/1053815111429968

Kalashnikova, M., Schwarz, I. C., & Burnham, D. (2016). OZI: Australian English communicative development inventory. *First Language, 36*(4), 407-427.

Kaplan, B. J., Wilson, B. N., Dewey, D., & Crawford, S. G. (1998). DCD may not be a discrete disorder. *Human Movement Science, 17*, 471-490.

Kern, S. (2007). Lexicon development in French-speaking infants. *First Language, 27*(3), 227-250. doi:10.1177/0142723706075789

Kern, S., & Gayraud, F. (2007). Influence of preterm birth on early lexical and grammatical acquisition. *First Language, 27*(2), 159-173.

King-Dowling, S., Missiuna, C., Rodriguez, M. C., Greenway, M., & Cairney, J. (2015). Co-occurring motor, language and emotional–behavioral problems in children 3–6 years of age. *Human Movement Science, 39*, 101-108. doi:10.1016/j.humov.2014.10.010

Klee, T., Carson, D. K., Gavin, W. J., Hall, L., Kent, A., & Reece, S. (1998). Concurrent and predictive validity of an early language screening program. *Journal of Speech, Language & Hearing Research, 41*(3), 627-641.

Klee, T., & Harrison, C. (2001). CDI words and sentences: Validity and preliminary norms for British English (supplement). Paper presented at the child language seminar, University of Hertfordshire. University of Canterbury research repository: http://hdl.handle.net/10092/3525.

Klee, T., Pearce, K., & Carson, D. K. (2000). Improving the positive predictive value of screening for developmental language disorder. *Journal of Speech, Language & Hearing Research, 43*(4), 821-833.

Korpilahti, P., Kaljonen, A., & Jansson-Verkasalo, E. (2016). Identification of biological and environmental risk factors for language delay: The let's talk STEPS study. *Infant Behavior and Development, 42*, 27-35. doi:http://dx.doi.org.proxy.library.lincoln.ac.uk/10.1016/j.infbeh.2015.08.008

Kreisman, D. (2012). The source of Black–White inequality in early language acquisition: Evidence from early head start. *Social Science Research, 41*(6), 1429-1450. doi:10.1016/j.ssresearch.2012.05.010

Kuhn, L. J., Willoughby, M. T., Wilbourn, M. P., Vernon-Feagans, L., Blair, C. B., & The Family Life Project Key Investigators. (2014). Early communicative gestures prospectively predict language development and executive function in early childhood. *Child Development, 85*(5), 1898-1914. doi:10.1111/cdev.12249

Laakso, M., Poikkeus, A., Katajamäki, J., & Lyytinen, P. (1999). Early intentional communication as a predictor of language development in young toddlers. *First Language, 19*(56), 207-231.

Law, J., Boyle, J., Harris, F., Harkness, A., & Nye, C. (1998). Screening for primary speech and language delay: A systematic review of the literature. *International Journal of Language & Communication Disorders, 33*, 21-23.

Law, J., Boyle, J., Harris, F., Harkness, A., & Nye, C. (2000a). The feasibility of universal screening for primary speech and language delay: Findings from a systematic review of the literature. *Developmental Medicine & Child Neurology, 42*(3), 190-200.

Law, J., Boyle, J., Harris, F., Harkness, A., & Nye, C. (2000b). Prevalence and natural history of primary speech and language delay: Findings from a systematic review of the literature. *International Journal of Language & Communication Disorders, 35*(2), 165-188. doi:10.1080/136828200247133

Law, J., Garrett, Z., & Nye, C. (2004). The efficacy of treatment for children with developmental speech and language delay/disorder: A meta-analysis. *Journal of Speech, Language & Hearing Research, 47*(4), 924-943.

Law, J., & Roy, P. (2008). Parental report of infant language skills: A review of the development and application of the communicative development inventories. *Child and Adolescent Mental Health, 13*(4), 198-206. doi:10.1111/j.1475-3588.2008.00503.x

Lee, J. (2011). Size matters: Early vocabulary as a predictor of language and literacy competence. *Applied Psycholinguistics, 32*(01), 69-92. doi:10.1017/S0142716410000299

Lenroot, R. K., Gogtay, N., Greenstein, D. K., Wells, E. M., Wallace, G. L., Clasen, L. S., . . . Giedd, J. N. (2007). Sexual dimorphism of brain developmental trajectories during childhood and adolescence. *NeuroImage, 36*, 1065-1073. doi:10.1016/j.neuroimage.2007.03.053

Leonard, H. C., & Hill, E. L. (2014). Review: The impact of motor development on typical and atypical social cognition and language: A systematic review. *Child and Adolescent Mental Health, 19*(3), 163-170.

Longobardi, E., Rossi-Arnaud, C., & Spataro, P. (2011). A longitudinal examination of early communicative development: Evidence from a parent-report questionnaire. *British Journal of Developmental Psychology, 29*(3), 572-593.

Longobardi, E., Spataro, P., & Rossi-Arnaud, C. (2014). The relationship between motor development, gestures and language production in the second year of life: A mediational analysis. *Infant Behavior and Development, 37*(1), 1-4. doi:http://dx.doi.org.proxy.library.lincoln.ac.uk/10.1016/j.infbeh.2013.10.002

Lueke, C., Ritterfeld, U., Grimminger, A., Rohlfing, K. J., & Liszkowski, U. (2015). (2015). Development of pointing gestures in children with and without primary language delay. Paper presented at the *Child Language Symposium,* University of Warwick. 118-118.

Lyytinen, H., Ahonen, T., Eklund, K., Guttorm, T. K., Laakso, M., Leinonen, S., . . . Puolakanaho, A. (2001). Developmental pathways of children with and without familial risk for dyslexia during the first years of life. *Developmental Neuropsychology, 20*(2), 535-554.

Lyytinen, P., Eklund, K., & Lyytinen, H. (2005). Language development and literacy skills in late-talking toddlers with and without familial risk for dyslexia. *Annals of Dyslexia, 55*(2), 166-192.

Lyytinen, P., Laasko, M. L., Poikkeus, A., & Rita, N. (1999). The development and predictive relations of play and language across the second year. *Scandinavian Journal of Psychology, 40*(3), 177-186.

Lyytinen, P., & Poikkeus, A. (1996). Continuity in early language development. *Infant Behavior and Development, 19 Supplement 1:594*, 594. doi:10.1016/S0163-6383(96)90648-7

Lyytinen, P., Poikkeus, A., Laakso, M., Eklund, K., & Lyytinen, H. (2001). Language development and symbolic play in children with and without familial risk for dyslexia. *Journal of Speech, Language & Hearing Research, 44*(4), 873-885.

Lyytinen, P., Poikkeus, A., Leiwo, M., Ahonen, T., & Lyytinen, H. (1996). Parents as informants of their child's vocal and early language development. *Early Child Development and Care, 126*(1), 15-25.

Mancilla-Martinez, J., & Vagh, S. B. (2013). Growth in toddlers' Spanish, English, and conceptual vocabulary knowledge. *Early Childhood Research Quarterly, 28*(3), 555-567. doi:http://dx.doi.org/10.1016/j.ecresq.2013.03.004

Mandel, D. R., Jusczyk, P. W., & Pisoni, D. B. (1995). Infants' recognition of the sound patterns of their own names. *Psychological Science, 6*(5), 314-317.

Marchman, V. A., & Bates, E. (1994). Continuity in lexical and morphological development: A test of the critical mass hypothesis. *Journal of Child Language, 21*(2), 339-366.

Mareschal, D., Johnson, M. H., Sirois, S., Spratling, M. W., Thomas, M. S. C., & Westerman, G. (2007). *Neuroconstructivism: How the brain constructs cognition (vol. I)*. New York, NY: Oxford University Press.

Marschik, P. B., Einspieler, C., Garzarolli, B., & Prechtl, H. F. R. (2007). Events at early development: Are they associated with early word production and neurodevelopmental abilities at the preschool age? *Early Human Development, 83*(2), 107-114.

Maw, R., Wilks, J., Harvey, I., Peters, T. J., & Golding, J. (1999). Early surgery compared with watchful waiting for glue ear and effect on language development in preschool children: A randomised trial. *Lancet, 353*(9157), 960-963.

McCall, R. B., Eichorn, D. H., Hogarty, P. S., Uzgiris, I. C., & Schaefer, E. S. (1977). Transitions in early mental development. *Monographs of the Society for Research in Child Development, 42*(3), 1-94. doi:10.2307/1165992

McClimans, L., Browne, J., & Cano, S. (2017). Clinical outcome measurement: Models, theory, psychometrics and practice. *Studies in History and Philosophy of Science,65*(66), 67-73.

McGregor, K. K., Sheng, L., & Smith, B. (2005). The precocious two-year-old: Status of the lexicon and links to the grammar. *Journal of Child Language, 32*(3), 563-583.

MEDCALC. (2017). Diagnostic test evaluation calculator. Retrieved from https://www.medcalc.org/calc/diagnostic_test.php

Meints, K., & Fletcher, K. (2011). Lincoln University Babylab toddler communicative development inventory, University of Lincoln, UK (version 2001). Retrieved from http://www.lincoln.ac.uk/home/psychology/research/lincolnbabylab/

Meints, K., Plunkett, K., Harris, P. L., & Dimmock, D. (2002). What is 'on' and 'under' for 15-, 18- and 24-month-olds? Typicality effects in early comprehension of spatial prepositions. *British Journal of Developmental Psychology, 20*(1), 113.

Meints, K., Plunkett, K., Harris, P. L., & Dimmock, D. (2004). The cow on the high street: Effects of background context on early naming. *Cognitive Development, 19*(3), 275-290. doi:http://dx.doi.org/10.1016/j.cogdev.2004.03.004

Meints, K., Plunkett, K., & Harris, P. L. (1999). When does an ostrich become a bird? The role of typicality in early word learning. *Developmental Psychology, 35*, 1072-1078. doi:10.1037/0012-1649.35.4.1072

Miller, S. E., & Marcovitch, S. (2015). Examining executive function in the second year of life: Coherence, stability, and relations to joint attention and language. *Developmental Psychology, 51*(1), 101-114.

Mills, D. L., Coffey-Corina, S., & Neville, H. J. (1997). Language comprehension and cerebral specialization from 13 to 20 months. *Developmental Neuropsychology, 13*(3), 397-445. doi:10.1080/87565649709540685

Miniscalco, C., Rudling, M., Råstam, M., Gillberg, C., & Åsberg, J. J. (2014). Imitation (rather than core language) predicts pragmatic development in young children with ASD: A preliminary longitudinal study using CDI parental reports...communicative developmental inventory. *International Journal of Language & Communication Disorders, 49*(3), 369-376.

Monson, B. B., Brownell, A., Inder, T. E., Warfield, S. K., Eaton-Rosen, Z., Kapur, K., ... Rogers, C. E. (2018). Differential rates of perinatal maturation of human primary and nonprimary auditory cortex. *ENeuro*, *5*(1). doi: 10.1523/ENEURO.0380-17.2017

Nash, H. M., Hulme, C., Gooch, D., & Snowling, M. J. (2013). Preschool language profiles of children at family risk of dyslexia: Continuities with specific language impairment. *Journal of Child Psychology and Psychiatry, 54*(9), 958-969.

Neuman, S.B., & Celano, D. (2001). Access to Print in Low-Income and Middle-Income Communities: An Ecological Study of Four Neighborhoods. *Reading Research Quarterly*, (36), 8-26.

NHS. (2018). Find digital tools to help you manage and improve your health. Retrieved from https://apps.beta.nhs.uk/

NICHD Early Child Care Research Network. (2005). Duration and developmental timing of poverty and children's cognitive and social development from birth through third grade. *Child Development, 76*(4), 795-810.

Nicholas, J. G., & Geers, A. E. (2008). Expected test scores for preschoolers with a cochlear implant who use spoken language. *American Journal of Speech-Language Pathology, 17*(2), 121-138 18p.

Norbury, C. F., Gooch, D. W. C., Baird, G., Charman, T., Simonoff, E., Vamvakas, G., & Pickles, A. (2016). The impact of nonverbal ability on prevalence and clinical presentation of language disorder: Evidence from a population study. *Journal of Child Psychology and Psychiatry, 57*(11), 1247-1258.

Oliver, B., Dale, P. S., Saudino, K. J., Petrill, S. A., Pike, A., & Plomin, R. (2002). The validity of a parent-based assessment of the nonverbal cognitive abilities of three-year-olds. *Early Child Development and Care, 172*, 337-348.

Orr, E., & Geva, R. (2015). Symbolic play and language development. *Infant Behavior & Development, 38*, 147-161. doi:10.1016/j.infbeh.2015.01.002

Oshima-Takane, Y., Goodz, E., & Derevensky, J. L. (1996). Birth order effects on early language development: Do secondborn children learn from overheard speech? *Child Development, 67*(2), 621-634.

Paavola, L., Kunnari, S., & Moilanen, I. (2005). Maternal responsiveness and infant intentional communication: Implications for the early communicative and linguistic development. *Child: Care, Health & Development, 31*(6), 727-735.

Pallant, J. (2013). *SPSS survival manual a step by step guide to data analysis using IBM SPSS* (5th ed.). Maidenhead: Open University Press/McGraw-Hill.

Pan, A. B., Rowe, M. L., Spier, E., & Tamis-LeMonda, C. (2004). Measuring productive vocabulary of toddlers in low-income families: Concurrent and predictive validity of three sources of data. *Journal of Child Language, 31*(03), 587-608. doi:10.1017/S0305000904006270

Paul, R. (2000). Predicting outcomes of early expressive language delay: Ethical implications. In D. V. M. Bishop, & L. B. Leonard (Eds.), *Speech and language impairments in children* (pp. 195-210). Hove: Psychology Press.

Pennington, B.F. (2006). From single to multiple deficit models of developmental disorders. *Cognition, 101*(2), 385–413.

Pérez-Pereira, M., Fernández, P., Gómez-Taibo, M. L., & Resches, M. (2014). Language development of low risk preterm infants up to the age of 30 months. *Early Human Development, 90*(10), 649-657.

Pérez-Pereira, M., & Resches, M. (2011). Concurrent and predictive validity of the Galician CDI. *Journal of Child Language, 38*(1), 121-140.

Pesco, D., & O'Neill, D. K. (2012). Predicting later language outcomes from the language use inventory. *Journal of Speech, Language & Hearing Research, 55*(2), 421-434.

Pickles, A., Simonoff, E., Conti-Ramsden, G., Falcaro, M., Simkin, Z., Charman, T., . . . Baird, G. (2009). Loss of language in early development of Autism and Specific Language Impairment. *Journal of Child Psychology & Psychiatry, 50*(7), 843-852. doi:10.1111/j.1469-7610.2008.02032.x

Pine, J. M. (1995). Variation in vocabulary development as a function of birth order. *Child Development, 66*(1), 272-281. doi:10.1111/1467-8624.ep9503233368

Prior, M., Bavin, E. L., Cini, E., Reilly, S., Bretherton, L., Wake, M., & Eadie, P. (2008). Influences on communicative development at 24 months of age: Child temperament, behaviour problems, and maternal factors. *Infant Behavior & Development, 31*(2), 270-279.

Pogash, C. (2016, March 31). In San Jose, poor find doors to library closed. *New York Times*, p. 13.

Putnick, D. L., Bornstein, M. H., Eryigit-Madzwamuse, S., & Wolke, D. (2017). Long-term stability of language performance in very preterm, moderate-late preterm, and term children. *The Journal of Pediatrics, 181*, 74-79.e3. doi:http://dx.doi.org.proxy.library.lincoln.ac.uk/10.1016/j.jpeds.2016.09.006

Reese, E., & Read, S. (2000). Predictive validity of the New Zealand MacArthur communicative development inventory: Words and sentences. *Journal of Child Language, 27*(2), 255-266.

Reilly, S., Wake, M., Bavin, E. L., Prior, M., Williams, J., Bretherton, L., . . . Ukoumunne, O. C. (2007). Predicting language at 2 years of age: A prospective community study. *Pediatrics, 120*(6), e1441-e1449.

Rescorla, L. A. (1989). The language development survey: A screening tool for delayed language in toddlers. *Journal of Speech and Hearing Disorders, 54*(4), 587-599.

Rescorla, L. A. (2002). Language and reading outcomes to age 9 in late-talking toddlers. *Journal of Speech, Language & Hearing Research, 45*(2), 360-371.

Rescorla, L. A. (2009). Age 17 language and reading outcomes in late-talking toddlers: Support for a dimensional perspective on language delay. *Journal of Speech, Language & Hearing Research, 52*(1), 16-31.

Rescorla, L. A. (2011). Late talkers: Do good predictors of outcome exist? *Developmental Disabilities, 17*(2), 141-150.

Rescorla, L. A. (2013). Late-talking toddlers: A 15-year follow-up. In L. A. Recorla, & P. S. Dale (Eds.), *Late talkers* (pp. 219-239)

Rescorla, L. A., Dahlsgaard, K., & Roberts, J. E. (2000). Late-talking toddlers: MLU and IPSyn outcomes at 3;0 and 4;0. *Journal of Child Language, 27*(3), 643-664.

Rescorla, L. A., & Dale, P. S. (2013). Where do we stand now? Conclusions and future directions. In L. A. Rescorla, & P. S. Dale (Eds.), *Late talkers language development, interventions, and outcomes* (pp. 377-387). London: Paul H. Brookes Publishing.

Rescorla, L. A., Roberts, J., & Dahlsgaard, K. (1997). Late talkers at 2: Outcome at age 3. *Journal of Speech, Language & Hearing Research, 40*(3), 556-567.

Rescorla, L. A., & Schwartz, E. (1990). Outcome of toddlers with specific expressive language delay. *Applied Psycholinguistics, 11*(4), 393-407.

Reznick, J. S., & Goldfield, B. A. (1992). Rapid change in lexical development in comprehension and production. *Developmental Psychology, 28*(3), 406-414.

Reznick, J. S., & Goldfield, B. A. (1994). Diary vs. representative checklist assessment of productive vocabulary. *Journal of Child Language, 21*(2), 465-472.

Roberts, J. E., Burchinal, M., & Durham, M. (1999). Parents' report of vocabulary and grammatical development of African American preschoolers: Child and environmental associations. *Child Development, 70*(1), 92-106.

Roberts, J. E., Burchinal, M. R., & Zeisel, S. A. (2002). Otitis media in early childhood in relation to children's school-age language and academic skills. *Pediatrics, 110*(4), 696-706.

Roberts, J. E., Rosenfield, R. M., & Zeisel, S. A. (2004). Otitis media and speech and language: A meta-analysis of prospective studies. *Pediatrics, 113*(3), 238-248.

Roberts, M. Y., & Kaiser, A. P. (2011). The effectiveness of parent-implemented language interventions: A meta-analysis. *American Journal of Speech-Language Pathology, 20*(3), 180-199. doi:10.1044/1058-0360(2011/10-0055)

Roberts, M. Y., & Kaiser, A. P. (2015). Early intervention for toddlers with language delays: A randomized controlled trial. *Pediatrics, 135*(4), 686-693. doi:10.1542/peds.2014-2134

Rodriguez, E. T., Tamis-LeMonda, C. S., Spellmann, M. E., Pan, B. A., Raikes, H., Lugo-Gil, J., & Luze, G. (2009). The formative role of home literacy experiences across the first three years of life in children from low-income families. *Journal of Applied Developmental Psychology, 30*(6), 677-695.

Roos, E., & Ellis Weismer, S. (2008). Language outcomes of late talking toddlers at preschool and beyond. *Perspectives on Language Learning and Education, 15*(3), 119-126.

Rose, S. A., Feldman, J. F., & Jankowski, J. J. (2009). A cognitive approach to the development of early language. *Child Development, 80*(1), 134-150.

Rowe, M. L., Özçalişkan, Ş., & Goldin-Meadow, S. (2008). Learning words by hand: Gesture's role in predicting vocabulary development. *First Language, 28*(2), 182-199.

Rowe, M. L., Raudenbush, S. W., & Goldin-Meadow, S. (2012). The pace of vocabulary growth helps predict later vocabulary skill. *Child Development, 83*(2), 508-525.

Ryan, A., Gibbon, F. E., & O'shea, A. (2016). Expressive and receptive language skills in preschool children from a socially disadvantaged area. *International Journal of Speech-Language Pathology, 18*(1), 41-52. doi:10.3109/17549507.2015.1089935

Sachse, S., Pecha, A., & von Suchodoletz, W. (2007). Early identification of developmental language disorders. Is ELFRA-2 (the German version of CDI) appropriate for general language screening at the age of 24 months? [Früherkennung von Sprachentwicklungsstörungen Ist der ELFRA-2 für einen generellen Einsatz bei der U7 zu empfehlen?] *Monatsschrift Kinderheilkunde, 155*(2), 140-145. doi:10.1007/s00112-006-1314-7

Sachse, S., Saracino, M., & von Suchodoletz, W. (2007). Early identification of developmental language disorders –Predictive value of the German version of the CDI – infant form (ELFRA-1). [Prognostische Validität des ELFRA-1 bei der Früherkennung von Sprachentwicklungsstörungen] *Klinische Pädiatrie, 219*, 17-22.

Sansavini, A., Guarini, A., Alessandroni, R., Faldella, G., Giovanelli, G., & Salvioli, G. (2007). Are early grammatical and phonological working memory abilities affected by preterm birth? *Journal of Communication Disorders, 40*(3), 239-256.

Sansavini, A., Guarini, A., Savini, S., Broccoli, S., Justice, L., Alessandroni, R., & Faldella, G. (2011). Longitudinal trajectories of gestural and linguistic abilities in very preterm infants in the second year of life. *Neuropsychologia, 49*(13), 3677-3689.

Scarborough, H. S. (1990). Very early language deficits in dyslexic children. *Child Development,* (6), 1728-1743.

Scarborough, H. S., & Dobrich, W. (1990). Development of children with early delay. *Journal of Speech and Hearing Research, 33*, 70-83.

Schaadt, G., Mannel, C., Meer, E., Pannekamp, A., & Friederici, A. D. (2016). Facial speech gestures: The relation between visual speech processing, phonological awareness, and developmental dyslexia in 10-year-olds. *Developmental Science, 19*(6), 1020-1034. doi:10.1111/desc.12346

Semel, E., Wiig, E. H., & Secord, W. (2000). *Clinical evaluation of language fundamentals (CELF-III UK)*. London: Psychological Corporation.

Serrat, E., Sanz-Torrent, M., Badia, I., Aguilar, E., Olmo, R., Lara, F., . . . Serra, M. (2010). The relationship between lexical learning and grammatical development. *Infancia y Aprendizaje, 33*(4), 435-448.

Sharkins, K. A., Leger, S. E., & Ernest, J. M. (2017). Examining effects of poverty, maternal depression, and children's self-regulation abilities on the development of language and cognition in early childhood: An early head start perspective. *Early Childhood Education Journal, 45*(4), 493-499.

Shriberg, L. D., Friel-Patti, S., Flipsen, P., & Brown, R. L. (2000). Otitis media, fluctuant hearing loss, and speech-language outcomes: A preliminary structural equation model. *Journal of Speech, Language & Hearing Research, 43*(1), 100-120.

Silva, C., Cadime, I., Ribeiro, I., Acosta, V., Lima, R., & Viana, F. L. (2017). Communicative development of Portuguese infants aged between 8 and 15 months. *Revista De Logopedia, Foniatria y Audologia, 37*(3), 121-129.

Silva, P. A., Williams, S., & McGee, R. (1987). A longitudinal study of children with developmental language delay at age three: Later intelligence, reading and behaviour problems. *Developmental Medicine and Child Neurology, 29*(5), 630-640.

Simonsen, H. G., Kristoffersen, K. E., Bleses, D., Wehberg, S., & Jørgensen, R. N. (2014). The Norwegian communicative development inventories: Reliability, main developmental trends and gender differences. *First Language, 34*(1), 3-23.

Singer, S., Corley, R., Guiffrida, C., & Plomin, R. (1984). The development and validation of a test battery to measure differentiated cognitive abilities in three-year-old children. *Educational and Psychological Measurement, 44*, 703-712.

Snowling, M. J., Stothard, S. E., Clarke, P., Bowyer-Crane, C., Harrington, A., Truelove, E., . . . Hulme, C. (2009). *York assessment of reading for comprehension*. London: GL Assessment.

Soriano-Ferrer, M., & Piedra Martínez, E. (2017). Review article: A review of the neurobiological basis of dyslexia in the adult population. *Neurología (English Edition), 32*(1), 50-58.

Southgate, V., & Meints, K. (2000). Typicality, naming and category membership in young children. *Cognitive Linguistics, Special Issue: Language Acquisition, 11*(1/2), 1-12.

Spencer, S., Clegg, J., & Stackhouse, J. (2012). Language and disadvantage: A comparison of the language abilities of adolescents from two different socioeconomic areas. *International Journal of Language & Communication Disorders, 47*(3), 274-285.

Spinelli, M., Fasolo, M., Shah, P., Genovese, G., & Aureli, T. (in press). The influence of early temperament on language development: The moderating role of maternal input. *Frontiers in Psychology.* doi: 10.3389/fpsyg.2018.01527

Spittle, A. J., Walsh, J. M., Potter, C., Mcinnes, E., Olsen, J. E., Lee, K. J., . . . Cheong, J. L. Y. (2017). Neurobehaviour at term-equivalent age and neurodevelopmental outcomes at 2 years in infants born moderate-to-late preterm. *Developmental Medicine & Child Neurology, 59*(2), 207-216.

Spitz, R. V., Tallal, P., Flax, J., & Benasich, A. A. (1997). Look who's talking: A prospective study of familial transmission of language impairments. *Journal of Speech, Language & Hearing Research, 40*(5), 990-1002.

Squires, J., Twombley, E., & Bricker, D. (2009). *ASQ-3 User's guide*. Baltimore: Paul H. Brookes Publishing Co.

Stanovich, K.E. (1988). Explaining the differences between the dyslexic and the garden-variety poor reader: the phonological-core variable-difference model. *Journal of Learning Disabilities, 21*(10), 590–604.

St Clair, M. C., Pickles, A., Durkin, K., & Conti-Ramsden, G. (2011). A longitudinal study of behavioral, emotional and social difficulties in individuals with a history of Specific Language Impairment (SLI). *Journal of Communication Disorders, 44*, 186-199. doi:10.1016/j.jcomdis.2010.09.004

Stevens, J. (1996). *Applied multivariate statistics for the social sciences* (3rd ed.). Mahwah, NJ: Lawrence Erlbaum.

Stiles, J. (1994). On the nature of informant judgments in inventory measures: ... and so what is it you want to know? *Monographs of the Society for Research in Child Development, 59*(5, Serial No. 242), 180-185.

Stolt, S., Haataja, L., Lapinleimu, H., & Lehtonen, L. (2009). The early lexical development and its predictive value to language skills at 2 years in very-low-birth-weight children. *Journal of Communication Disorders, 42*(2), 107-123. doi:http://dx.doi.org/10.1016/j.jcomdis.2008.10.002

Stolt, S., Klippi, A., Launonen, K., Munck, P., Lehtonen, L., Lapinleimu, H., & Haataja, L. (2007). Size and composition of the lexicon in prematurely born very-low-birth-weight and full-term Finnish children at two years of age. *Journal of Child Language, 34*(2), 283-310.

Stolt, S., Lind, A., Matomäki, J., Haataja, L., Lapinleimu, H., & Lehtonen, L. (2016). Do the early development of gestures and receptive and expressive language predict language skills at 5;0 in prematurely born very-low-birth-weight children? *Journal of Communication Disorders, 61*, 16-28. doi:10.1016/j.jcomdis.2016.03.002

Styles, S., & Plunkett, K. (2009). What is 'word understanding' for the parent of a one-year-old? Matching the difficulty of a lexical comprehension task to parental CDI report. *Journal of Child Language, 36*(4), 895-908.

Suizzo, M. A., & Bornstein, M. H. (2006). French and European American child-mother play: Culture and gender considerations. *International Journal of Behavioral Development, 30*(6), 498-508. doi:10.1177/0165025406071912

Sundqvist, A., Nordqvist, E., Koch, F. -., & Heimann, M. (2016). Early declarative memory predicts productive language: A longitudinal study of deferred imitation and communication at 9 and 16 months. *Journal of Experimental Child Psychology, 151*, 109-120.

Suttora, C., & Salerni, N. (2012). Gestural development and its relation to language acquisition in very preterm children. *Infant Behavior & Development, 35*(3), 429-438. doi:10.1016/j.infbeh.2012.02.008

Syrnyk, C., & Meints, K. (2017). Bye-bye mummy - word comprehension in 9-month-old infants. *British Journal of Developmental Psychology, 35*, 202-217.

Szagun, G., Steinbrink, C., Franik, M., & Stumper, B. (2006). Development of vocabulary and grammar in young German-speaking children assessed with a German language development inventory. *First Language, 26*(3), 259-280.

Szagun, G., Stumper, B., & Schramm, S. (2009). *FRAKIS Fragebogen zur Frühkindlichen Sprachentwicklung: FRAKIS (Standardform), FRAKIS-K (Kurzform)*. Frankfurt am Main: Pearson. Retrieved from http://proxy.library.lincoln.ac.uk/login?url=http://search.ebscohost.com/login.aspx?direct=true&db=edsoai&AN=edsoai.761416068&site=eds-live&scope=site; http://www.egms.de/en/meetings/dgpp2010/10dgpp61.shtml

Tabachnick, B. G., & Fidell, L. S. (2001). *Using multivariate statistics* (4th ed. ed.). Needham Heights, MA: Allyn & Bacon.

Tartter, V. C. (1998). *Language processing in atypical populations*. Thousand Oaks: SAGE.

Templeton, G. F. (2011). A two-step approach for transforming continuous variables to normal: Implications and recommendations for IS research. *Communications of the Association for Information Systems, 28*, 41-58.

Thal, D. J., Bates, E., Goodman, J., & Jahn-Samilo, J. (1997). Continuity of language abilities: An exploratory study of late- and early-talking toddlers. *Developmental Neuropsychology, 13*(3), 239-273.

Thal, D. J., Bates, E., Zappia, M. J., & Oroz, M. (1996). Ties between lexical and grammatical development: Evidence from early-talkers. *Journal of Child Language, 23*(2), 349-368.

Thal, D. J., Marchman, V. A., & Tomblin, J. B. (2013). Late-talking toddlers: Characterization and prediction of continuing delay. In L. A. Rescorla, & P. S. Dale (Eds.), *Late talkers: Language development, interventions and outcomes* (pp. 169-201). Baltimore, MD: Brookes.

Thal, D. J., O'Hanlon, L., Clemmons, M., & Fralin, L. (1999). Validity of a parent report measure of vocabulary and syntax for preschool children with language impairment. *Journal of Speech, Language, and Hearing Research, 42*, 482-496. doi:10.1044/jslhr.4202.482

Thal, D. J., Swaine, K., Harrison, D., & Matt, G. (1998). Prediction of language delay at 16 months of age using parent report. *Infant Behavior and Development, 21, Supplement*, 717. doi:http://dx.doi.org/10.1016/S0163-6383(98)91930-0

Thomas, M. S. C., Annaz, D., Ansari, D., Scerif, G., Jarrold, C., & Karmiloff-Smith, A. (2009). Using developmental trajectories to understand developmental disorders. *Journal of Speech, Language, and Hearing Research, 52*(2), 336-359.

Tincoff, R., & Juscyzk, P. W. (1999). Some beginnings of word comprehension in 6-month-olds. *Psychological Science, 10*(2), 172.

Todd, B. K., Thommessen, S. A. O., & Barry, J. A. (2016). Preferences for 'gender-typed' toys in boys and girls aged 9 to 32 months. *Infant and Child Development,* doi:10.1002/icd.1986

Tomasello, M. J., & Mervis, C. B. (1994). The instrument is great, but measuring comprehension is still a problem. *Monographs of the Society for Research in Child Development, 59*(5, Serial No. 242), 174-179.

Tomblin, J. B. (2012). Children with Specific Language Impairment. In E. L. Bavin (Ed.), *The Cambridge handbook of child language* (2nd ed., pp. 417-431)

Tomblin, J. B., Records, N. L., Buckwalter, P., Zhang, X., Smith, E., & O'Brien, M. (1997). Prevalence of Specific Language Impairment in kindergarten children. *Journal of Speech, Language & Hearing Research, 40*(6), 1245.

Tomblin, J. B., Smith, E., & Zhang, X. (1997). Epidemiology of Specific Language Impairment: Prenatal and perinatal risk factors. *Journal of Communication Disorders, 30*, 325-344.

Tomblin, J. B., Zhang, X., Buckwalter, P., & O'Brien, M. (2003). The stability of primary language disorder: Four years after kindergarten diagnosis. *Journal of Speech, Language, and Hearing Research, 46*, 1283-1296.

Torppa, M., Tolvanen, A., Poikkeus, A. M., Eklund, K., Lerkkanen, M. K., Leskinen, E., & Lyytinen, H. (2007). Reading development subtypes and their early characteristics. *Annals of Dyslexia, 57*(1), 3–52.

Tucker-Drob, E., & Briley, D. A. (2014). Continuity of genetic and environmental influences on cognition across the life span: A meta-analysis of longitudinal twin and adoption studies. *Psychological Bulletin, 140*(4), 949-979. doi:10.1037/a0035893; 10.1037/a0035893.supp (Supplemental)

Ullrich, K., & von Suchodoletz, W. (2011). Früherkennung von Sprachentwicklungsstörungen bei der U7 diagnostische Validität der Elternfragebögen SBE-2-KT und ELFRA-2 (German). *Monatsschrift Kinderheilkunde, 159*(5), 461-467.

Unhjem, A., Eklund, K., & Nergård-Nilssen, T. (2015). Early markers of language delay in children with and without family risk for dyslexia. *First Language, 35*(3), 254-272.

Vach, W., Bleses, D., & Jørgensen, R. (2010). Construction of a Danish CDI short form for language screening at the age of 36 months: Methodological considerations and results. *Clinical Linguistics & Phonology, 24*(8), 602-621.

van der Leij, A., Lyytinen, H., & Zwarts, F. (2001). The study of infant cognitive process in dyslexia. In A. J. Fawcett (Ed.), *Dyslexia: Theory and good practice* (pp. 160-181). London: Whurr.

van Zuijen, T. L., Plakas, A., Maassen, B. A. M., Maurits, N. M., & Leij, A. (2013). Infant ERPs separate children at risk of dyslexia who become good readers from those who become poor readers. *Developmental Science,* (4), 554-563.

Velikonja, T., Edbrooke-Childs, J., Calderon, A., Sleed, M., Brown, A., & Deighton, J. (2016). The psychometric properties of the ages & stages questionnaires for ages 2-2.5: A systematic review. *Child: Care, Health and Development, 43*(1), 1-17.

Veness, C., Prior, M., Eadie, P., Bavin, E., & Reilly, S. (2014). Predicting Autism diagnosis by 7 years of age using parent report of infant social communication skills. *Journal of Paediatrics and Child Health, 50*(9), 693-701.

Vernon-Feagans, L., Pancsofar, N., Willoughby, M., Odom, E., Quade, A., & Cox, M. (2008). Predictors of maternal language to infants during a picture book task in the home: Family SES, child characteristics and the parenting environment. *Journal of Applied Developmental Psychology, 29*(3), 213-227.

Viholainen, H., Ahonen, T., Cantell, M., Lyytinen, P., & Lyytinen, H. (2002). Development of early motor skills and language in children at risk for familial dyslexia. *Developmental Medicine & Child Neurology, 44*(11), 761-769. doi:10.1017/S0012162201002894

von Suchodoletz, W., Kademann, S. & Tippelt, S. (2009). Sprachbeurteilung durch Eltern: Kurztest für die U7a (SBE-3-KT). Klinik und Poliklinik für Kinder- und Jugendpsychiatrie, Psychotherapie und Psychosomatik, Ludwig-Maximilians-Universität München, Retrieved from www.kjp.med.uni-muenchen.de/sprachstoerungen/sprachentwicklung.php

von Suchodoletz, W., & Sachse, S. (2008). SBE-2-KT: Sprachbeurteilung durch Eltern – Kurztest für die U7. Klinik und Poliklinik für Kinder- und Jugendpsychiatrie, Psychotherapie und Psychosomatik, Ludwig-Maximilians-Universität München, Retrieved from www.kjp.med.uni-muenchen.de/sprachstoerungen/SBE-2-KT.php

Wagner, R. K., Torgesen, J. K., & Rashotte, C. A. (1999). *Comprehensive test of phonological processes*. Austin, TX: Pro-Ed.

Weber, A., Fernald, A., & Diop, Y. (2017). When cultural norms discourage talking to babies: Effectiveness of a parenting program in rural Senegal. *Child Development, 88*(5), 1513-1526. doi:10.1111/cdev.12882

Wechsler, D. (2002). *Wechsler preschool and primary scale of intelligence-third edition*. San Antonio, TX: The Psychological Corporation.

Wehberg, S., Vach, W., Bleses, D., Thomsen, P., Madsen, T. O., & Basbøll, H. (2007). Danish children's first words: Analysing longitudinal data based on monthly CDI parental reports. *First Language, 27*(4), 361-384.

Wehberg, S., Vach, W., Bleses, D., Thomsen, P., Madsen, T. O., & Basbøll, H. (2008). Girls talk about dolls and boys about cars? Analyses of group and individual variation in Danish children's first words. *First Language, 28*(1), 71-85. doi:10.1177/0142723707081729

Westerlund, M., Berglund, E., & Eriksson, M. (2006). Can severely language delayed 3-year-olds be identified at 18 months? Evaluation of a screening version of the MacArthur-Bates communicative development inventories. *Journal of Speech, Language & Hearing Research, 49*(2), 237-247.

Wetherby, A. M., Goldstein, H., Cleary, J., Allen, L., & Kublin, K. (2003). Early identification of children with communication disorders: Concurrent and predictive validity of the CSBS developmental profile. *Infants & Young Children: An Interdisciplinary Journal of Early Childhood Intervention, 16*(2), 161-174.

Whitehouse, A. J. O., Watt, H. J., Line, E. A., & Bishop, D. V. M. (2009). Adult psychosocial outcomes of children with specific language impairment, pragmatic language impairment and Autism. *International Journal of Language and Communication Disorders, 44*(4), 511-528.

Whitehurst, G. J., & Fischel, J. E. (1994). Practitioner Review: Early Developmental Language Delay: What, If Anything, Should the Clinician Do About It? *Journal of Child Psychology and Psychiatry, 35*(4), 613 – 648.

Wiig, E. H., Secord, W., & Semel, E. (1992). *Clinical evaluation of language fundamentals- preschool*. Psychological Corporation: San Antonio, TX.

Wiig, E. H., Secord, W., & Semel, E. (2004). *Clinical evaluation of language fundamentals, preschool* (2nd ed.). San Antonio, TX: The Psychological Corporation.

Williams, K. T. (1997). *Expressive vocabulary test*. Circle Pines, MN: American Guidance Service.

Williams, K. T. (2007). *Expressive vocabulary test, second edition*. Circle Pines, MN: American Guidance Service.

Yoder, P. J., Warren, S. F., & Biggar, H. A. (1997). Stability of maternal reports of lexical comprehension in very young children with developmental delays. *American Journal of Speech-Language Pathology, 6*, 59-64.

Yoder, P. J., Warren, S. F., & McCathren, R. B. (1998). Determining spoken language prognosis in children with developmental disabilities. *American Journal of Speech-Language Pathology, 7*, 77-87. doi:10.1044/1058-0360.0704.77

Zambrana, I. M., Pons, F., Eadie, P., & Ystrom, E. (2014). Trajectories of language delay from age 3 to 5: Persistence, recovery and late onset. *International Journal of Language & Communication Disorders, 49*(3), 304-316. doi:10.1111/1460-6984.12073

Zimmerman, I. L., Steiner, V. G., & Pond, R. E. (2011). *Preschool language scales, fifth edition (PLS-5)*. San Antonio, TX: The Psychological Corporation.

Zimmerman, I. L., Steiner, V. G., & Pond, R. E. (2002). *Preschool language scale - fourth edition*. San Antonio, TX: The Psychological Corporation.

Zimmerman, I. L., Steiner, V. G., & Pond, R. E. (2014). *PLS-5 UK preschool language scales - fifth edition - UK examiner's manual*. London: Pearson Assessment.

Zweig, M. H., & Campbell, G. (1993). Receiver operator characteristic (ROC) plots: A fundamental evaluation tool in clinical medicine. *Clinical Chemistry, 39*, 561-577.

Appendices

Appendix 1 – Details on literature search

Databases to be used:

1) Webinfo
2) Psychinfo (part of EBSCO database)
3) Webofscience
4) Sciencedirect
5) Psycharticles
6) Academic search complete
7) Medline

Keywords:

Keywords were combined across the different boxes, see example below. During the search, articles in all languages were accepted and authors were contacted if an English translation was also available. If articles were written in languages other than English or German, they had to be discarded for this review.

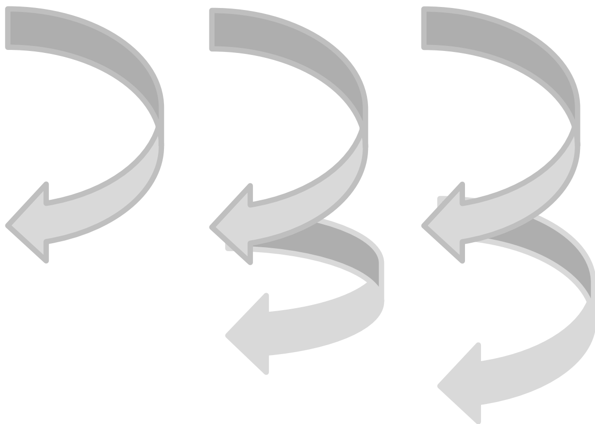
Example: Keyword combinations for
1) MacArthur* **AND OR** "Communicative Development Inventory" **OR** CDI

Search 1 | Search 2 | Search 3

1) MacArthur* **AND OR** "Communicative Development Inventory" **OR** CDI
2) Parent-report*

1) Predict* **AND** value **OR** validity **OR** language

1) Infan* **OR** Toddler
2) "Early development"



Database search for each keyword combination (6 combinations overall) and stop after 2-3 pages with no relevant info.

Appendix 2 - Selected information of studies used in literature review

Table 1. *Selected information of studies used in literature review*

Author(s)	Year	Language	type of participants	type of CDI (Long/Short Form = SF/LF)	norms	aim
Hsu & Iyer	2016	American English	representative sample of the area NICHD SECCYD data (varied SES, no cognitive deficits)	CDI:WG (LF)	yes	Are early gestures and vocabularies associated with the risk of developing later language impairment?
Thal et al.	2013	American English	population-based sample	CDI:WG (LF), CDI:WS (LF)	yes	predicting language delay and impairment from earlier CDI:WG scores (gesture, comprehension and production), studying differences between children depending on ability group
Friend et al.	2012	American English	Predominantly middle class, demographically diverse sample (includes children exposed to another language)	CDI:WG (LF), CDI:WS (LF)	yes (but only up to 18 months)	assessing the convergent and predictive validity of the CCT and CDI:WG
Kreisman et al.	2012	American English	White and Black low-income families, English speaking households, Head start intervention (EHSRE)	CDI:WG (SF, but including gesture scale) at 14m, CDI:WS (SF) at 24m	yes	it is being tested if a set of control variables can account for the differences in language development between Black and White children during the first 3 years of life, and if certain effects maybe stronger for one over the other population over time
Rose et al.	2009	American English, small % Spanish	partially preterm (89.8% monolingual English speakers, 7% bilingual Spanish, 3.2% solely Spanish)	CDI:WG (SF, but including gesture scale from LF)	yes	examines if language development is tied to language specific processes alone or if it is also dependent on other basic cognitive processes; assesses the predictive validity of the CDI(SF)
Fenson et al.	2007	American English	sub-dataset from US CDI norm sample	CDI:WG (LF), CDI:WS (LF)	yes	assessing the predictive validity of the CDI:WG

Fish & Pinkerman	2003	American English	low-SES, rural	CDI:WG (LF)	yes	investigating the effects of contextual (e.g. SES, maternal personality and behaviour, family literacy and academic expectations), child (CDI scores at 15 months, temperament at 4 years) and maternal interaction (at 9 months or 4 years) variables on children's language at 4 and 5;4 years of age using the PLS-3
Feldman et al.	2000	American English	sociodemgraphically diverse sample, healthy children, English spoken at home, children/families were excluded who exhibited criteria that could affect global or language developmental outcomes (e.g. low birthweight, small gestational age, multiple births, malformation, serious or chronic illness, social or intellectual disability etc.)	CDI:WG (LF), CDI:WS (LF)	yes	assessed the measurement properties of the CDI:WG and CDI:WS on a more sociodemographically diverse sample than the original norm sample
Baumwell et al.	1997	American English	middle to upper middle class households, firstborn, term, monolingual English children	CDI:WG (LF) (interview technique used)	maternal interviews using CDI and "Language and Gesture Inventory" and using coding system for these instruments	examining covariation among different maternal behaviors and the prediction of child language comprehension between 9 and 13 months
Thal et al.	1997	American English	Study 1: US norm sample, Study 2: mostly first-borns, ethnically and racially representative of the US, range of socio-economic backgrounds, not listening to another at a regular basis at study	CDI:WG (LF), CDI:WS (LF)	yes	assessing the continuity and stability of language of late and early talkers

			start but some were exposed to a second language due to child care arrangements			
Fenson et al.	1994	American English	sub-dataset from US CDI norm sample	CDI:WG (LF), CDI:WS (LF)	yes	assessing the predictive validity of the CDI:WG
Cochet & Byrne	2016	British English	no info	Oxford CDI for all age groups (no gesture scale)	no standardized norms for the UK	examined developmental stability between early social and communicative (including gesture, language) abilities between 11 and 41 months
Duff, Nation, Plunkett & Bishop	2015	British English	31% late talkers at 18 months	Oxford CDI (LF) (no onomatopoeia, no gesture scale)	no standardized norms for the UK	compare language and literacy outcomes at 8 years between children classified as average and late talkers at 18 months; language status at 4 years was available for a subsample it was hypothesised that language status was better at predicting later outcomes than at 18 months
Duff, Reen, Plunkett & Nation	2015	British English	above average SES, 46% had a first degree relative with language and/or reading difficulty	Oxford CDI (no gesture scale)	no standardized norms for the UK	Does infant vocab ability predict school-age reading and language skills? Does family risk contribute to the prediction?
Hamadani et al.	2010	Bengali	poor, rural	CDI:WG (SF)	no standardized norms for Bangladesh (may be suitable for other rural areas in Bangladesh)	development of a CDI:WG (SF) measure between 12 and 18 months in rural Bangladesh, test the concurrent and predictive validity and its sensitivity to other measures (e.g. stimulation, Bayley scores)
Bornstein et al.	2006	Dutch in Belgium	first-born, monolinguals, normally developing and term birth children	CDI:WG (LF, comp, prod, comp + prod cumulative score=p, comp, prod, comp + prod cumulative score), CDI:WS (LF)	yes	assessing the stability and continuity of the CDI across the second year for maternal reports but also comparing scores between reporters (usually mother, father, other carer) and a cumulative score

Bavin et al.	2008	English (Australia) Victoria Study	SES diverse sample (low SES underrepresented), healthy children with parents who spoke and understood English	CDI:WG, CDI:WS (LF) (at 8m: sole use of subsection "First Communicative Gestures, Games and Routines" except for "Pretending to be a Parent" and "Imitating Other Adult Actions")	no standardized norms for Australia, some vocabulary adaptations in both versions	examine the extent to which communicative behaviours at 8 and 12m could predict vocabulary at 12 and 24m
Korpilahtia et al.	2016	Finnish	population-based sample, monolingual Finnish, children had no severe developmental illnesses or impairments	CDI:WG (13m, vocab: comp, prod), CDI:WS (24m, vocab: prod)	yes (Lyytinen, 1999)	explore risk factors (biological, environmental) for language delay; impacts of fathers in language development; association between CDI vocab scores, language tests results and parents concerns with regards to language dev
Lyytinen et al.	2001	Finnish	half at risk for dyslexia, children from similar educational backgrounds	CDI:WG (LF)	no standardized norms for Finnish, questionnaire adapted in terms of cultural and linguistic relevance	developmental pathways up to 5 years were compared for children with and without familial risk of dyslexia
Lyytinen et al.	1999	Finnish	full-term children, half of them at risk for dyslexia but no differences in terms of Bayley mental (MDI) scores at 24m	CDI:WG (LF) (vocab comp and production scales used at 14m), CDI:WS (LF) (used at 18 (prod) and 24m (prod and MSL))	no standardized norms for Finnish, questionnaire adapted in terms of cultural and linguistic relevance	examine changes in children's nonsymbolic and sybolic play behaviour between 14 and 18 months, assess how early play and language predict language and cognitive skills at 24 m taking into account other factors (i.e. gender, parental education)

Laasko et al.	1999	Finnish	full-term children, native Finnish parents, representative parental educational levels, diverse parental reading skills from poor to average	CDI:WG (LF) (at 14m: comp, prod, sum score of 5 out of 6 subscales of "Actions amd Gestures", subscale of "Pretend objects" was excluded due to little variation in scores), CDI:WS (LF) (used at 24m (prod and MSL), sum score with Bayley expressive score)	no standardized norms for Finnish, questionnaire adapted in terms of cultural and linguistic relevance	investigate early intentional communication skills (actions and gestures, joint attention) as a predictor of language skills in toddlers
Lyytinen et al.	1996	Finnish	parents were Finnish natives, SES representative for the Finnish population, half the parents had problems with reading, children were full-term	CDI:WG (LF) (gestures: symbolic gestures used only), CDI:WS (LF) (grammar: use of suffixes used only)	no standardized norms for Finnish, questionnaire adapted in terms of cultural and linguistic relevance	examine the continuity of children's early vocalizations and language up to 18 months
Sachse et al.	2007	German	monolingual German families, adresses were taken from births notices in newspaper (61% at 12m, 95% at 18m)	ELFRA-1 (Infant CDI + LDS, comprehension: vocab checklist, reactions towards language; production: vocab checklist, production of sounds and language; fine motor skills, gestures), ELFRA-2 (vocabulary, syntax, morphology)	yes	examine the predictive validity of the ELFRA-1

Unhjem et al.	2015	Norwegian	at risk for dyslexia, typical, within typical range for cognitive ability, no neurological disabilities, monolingual Norwegian	CDI:WG (12m, 15m, 18m; only vocab checklist used) CDI:WS (18m, 24m; only vocab checklist used)	yes	explore the continuity of langue development during the second year in children at risk of dyslexia and control children; examine predictive relations between receptive and expressive from 12 to 24 months; investigate if at risk children were more likely to become late talkers at 24m and if knowing the risk status would help improve the prediction for late talking
Guiberson	2008	Spanish in US (Mexican immigrants)	mostly monolingual Spanish speaking parents (not more than 5% of the time), low income, low SES families	CDI:WG (words produced), CDI:WS (words produced)	yes (standardized for Mexico)	examine the concurrent the CDI (CDI:WG and CDI:WS) with observed lexical diversity measure (at 14-16 months and 30-32 months), predictive association of CDI:WG (at 14-16 months) with observed lexical diversity (at 30-32 months) and CDI:WS (at 30-32 months)
Westerlund et al.	2006	Swedish	Swedish primary language, healthy children	CDI:WG (SF)	yes	examine which component of the CDI (comp, prod, gest) best predicts severe language delay at 3 years (predictors were studied using ROC curves), traditional assessment outcomes were comapred with CDI results, it was investigated if other cutoff than 8 words would be better using the CDI

Appendix 3 - Differences in CDI vocabulary for preterm versus term children

Table 1. *Statistical analysis for CDI vocabulary scores between preterm and term children; Means and standard deviations*

Time at CDI completion	status	N	M	SD	df	t	p
12months							
1 Comprehension	term	7	78.57	76.84			
	preterm	7	51.29	13.33			
					7.1	0.33	0.751
2 Comprehension	term	7	46	33.15			
	preterm	7	51.29	13.33			
					12	-1	0.337
3 Comprehension	term	7	69.57	70.54			
	preterm	7	51.29	13.33			
					6.3	-0.66	0.535
1 Production	term	7	5	7.53			
	preterm	7	4.29	5.22			
					12	-0.03	0.974
2 Production	term	7	5.29	4.82			
	preterm	7	4.29	5.22			
					12	0.44	0.671
3 Production	term	7	5	5.86			
	preterm	7	4.29	5.22			
					12	0.29	0.78
18 months							
1 Comprehension	term	7	217.86	90.89			
	preterm	7	169.71	126.65			
					12	0.82	0.43
2 Comprehension	term	7	195.57	74.03			
	preterm	7	169.71	126.65			
					12	0.47	0.649
3 Comprehension	term	7	201	66.17			
	preterm	7	169.71	126.65			
					12	0.58	0.573
1 Production	term	7	76.57	94.28			
	preterm	7	48.57	34.36			
					12	0.43	0.672
2 Production	term	7	46.29	61.77			
	preterm	7	48.57	34.36			
					12	-0.89	0.392
3 Production	term	7	53.86	45.93			
	preterm	7	48.57	34.36			
					12	0.51	0.623
24 months							
1 Comprehension	term	7	442.71	130.11			

	preterm	7	418.43	191.67			
					12	0.28	0.786
2 Comprehension	term	7	442.86	149.73			
	preterm	7	418.43	191.67			
					12	0.27	0.795
3 Comprehension	term	7	420.14	108.01			
	preterm	7	418.43	191.67			
					12	0.02	0.984
1 Production	term	7	325.86	218.26			
	preterm	7	266.86	216.89			
					12	0.51	0.621
2 Production	term	7	263.29	200.93			
	preterm	7	266.86	216.89			
					12	-0.03	0.975
3 Production	term	7	299.43	128.18			
	preterm	7	266.86	216.89			
					12	0.34	0.738
36 months							
1 Production	term	7	74.86	17.38			
	preterm	7	59	35.59			
					7.21	0.34	0.743
2 Production	term	7	65.71	39.4			
	preterm	7	59	35.59			
					12	-0.23	0.824
3 Production	term	7	72.14	34.22			
	preterm	7	59	35.59			
					12	-0.35	0.732

* vocabulary scores were log transformed to be normal (i.e. comprehension at 12months and production 12 and 18months), Means and SD were displayed as Raw scores (i.e. not log transformed scores)

Appendix 4 – Family Questionnaire

Family Questionnaire

A baby's health and family life can have a big influence on the words and gestures they learn.
So we would like to ask a few questions about these.

Please feel free to leave out any questions that you don't want to answer.

Please do not write your name or address on any part of this questionnaire so that the information is anonymous and confidential. Please remember to fill in this questionnaire for the same child you are filling in the UK-CDI for.

A. YOUR CHILD'S HEALTH AND DEVELOPMENT

1. At what week of pregnancy was your child born?	Week 33 or before	O	Week 34 to 36	O	Week 37 or later	O
2. How much did your child weigh at birth?	Up to 5lb 8oz	O	5lb 9oz to 9lb 14oz	O	9lb 15oz or over	O

3. Has your child had an ear infection / glue ear for longer than 3 months, 4 to 6 ear infections within a 6 month period, or another identified hearing problem (e.g. at newborn hearing screening). Yes O No O

If yes please give details here:

4. Is there anyone in the child's immediate family (brothers/sisters/parents only) with a speech / language difficulty or dyslexia? Yes O No O

If yes please give details here:

5. Does your child have a developmental disability (e.g. Cerebral Palsy, ASD, Fragile X syndrome, Muscular dystrophy, Di George syndrome, Down's syndrome, Williams syndrome)? Yes O No O

If yes please give details here:

6. Does your child have a hearing or visual impairment? Yes O No O

If yes please give details here:

7. Have you or anyone else had any concerns about your child's hearing or communication? Yes O No O

If yes please give details here:

B. YOUR CHILD'S FAMILY

1. How many siblings does your child have? (include full and half siblings)	0	O	1	O	2	O	3	O	4 or more	O

2. Is this child the first child of his or her mum (i.e. does mum have any older children)?	Yes	O	No	O

If no, how many older children does mum have?	1	O	2	O	3 or more	O

3. Is your child a twin/multiple birth?	Yes	O	No	O

4. Is your child:

White British/Irish	O	Mixed Ethnicity: White and other	O	Asian/Asian British	O
Black/African/Caribbean /Black British	O	Other ethnic group (please give details):			

5. How long does your child spend at your home address?

Less than half the year	O	About half the year	O	More than half the year	O	All year	O

6. Which other people over 18 years old live in this home with you and your child?

Mum	O	Dad	O	Grandparent/s	O

Other related adults (please say how many)	0	O	1	O	2	O	3	O	more	O
Other non-related adults (please say how many)	0	O	1	O	2	O	3	O	more	O

7. How many other children live in your home with you? (please say how many in each age range)

Children 0-18 months	0	O	1	O	2	O	3	O	more	O
Children 19 months- 3 years 11 months	0	O	1	O	2	O	3	O	more	O
Children 4- 11 years	0	O	1	O	2	O	3	O	more	O

8. How many bedrooms are in your home?	1 O	2 O	3 O	4 O	5+ O

Please only answer questions 9-13 if your child lives at another home for part of the year

9. Please give postcode ____________________

10. How long do they spend at this other home address?

Less than half the year O About half the year O More than half the year O

11. Which other people over 18 years old live in this other home with your child?

Mum O Dad O Grandparent/s O

Other related adults (please say how many)	0 O	1 O	2 O	3 O	more O
Other non-related adults (please say how many)	0 O	1 O	2 O	3 O	more O

12. How many other children live in this other home? (please say how many in each age range)

Children 0- 18 months	0 O	1 O	2 O	3 O	more O
Children 19 months- 3 years 11 months	0 O	1 O	2 O	3 O	more O
Children 4- 11 years	0 O	1 O	2 O	3 O	more O
Children 12- 17 years	0 O	1 O	2 O	3 O	more O

13. How many bedrooms are in this other home?	1 O	2 O	3 O	4 O	5+ O

Child's Mum

14. Child's mum's age is….		Up to 20 years old O	21-25 years old O
26-30 years old O		31-35 years old O	36+ years old O
15. How old was child's mum when her oldest child was born?		Up to 20 years old O	21-25 years old O
26-30 years old O		31-35 years old O	36+ years old O

16. Mum is... Married/Civil Partnered O Living with partner O
Single O Separated/Divorced O Widowed O

17. Mum's highest education is... No formal qualifications O GCSE/O Level/NVQ Level 1 or 2/ similar O
A Level/NVQ Level 3/ similar O University degree/HND/HNC/NVQ Level 4 or 5/similar O Postgraduate/similar e.g. (PGCE, PhD, MA etc.) O

18. Mum's work status is... (current job or last paid job) Never worked, have only been in training or education O
An employee O Self-employed (with employees) O Self-employed (without employees) O

19. Mum's current/last job title: (please be specific)

20. How many people work for mum's employer or for mum if she is/was an employer? (only answer this question if mum is/was an employee or self-employed with employees)

0 O 1-24 O 25+ O

Child's Dad

21. Child's dad's age is... Up to 20 years old O 21-25 years old O
26-30 years old O 31-35 years old O 36+ years old O

22. Dad is... Married/Civil Partnered O Living with partner O
Single O Separated/Divorced O Widowed O

23. Dad's highest education is... No formal qualifications O GCSE/O Level/NVQ Level 1 or 2/ similar O
A Level/NVQ Level 3/ similar O University degree/HND/HNC/NVQ Level 4 or 5/similar O Postgraduate/similar e.g. (PGCE, PhD, MA etc.) O

24. Dad's work status is... (current job or last paid job) Never worked, have only been in training or education O
An employee O Self-employed (with employees) O Self-employed (without employees) O

25. Dad's current/last job title: (please be specific)

26. How many people work for dad's employer or for dad if he is/was an employer? (only answer this question if dad is/was an employee or self-employed with employees)

27. What is the overall household income (before tax) per year in your child's main home?

£0-£14000	O	£14,001-£24,000	O	£24,001-£42,000	O	£42,001 or more	O

C. ALL ABOUT YOUR CHILD'S DAY

1. Who looks after your child? Please tell us about everyone who looks after your child for at least half a day in a typical week.

Child's mum	O	Child's dad	O	Other carer or carers	O

2. If you have told us about other carers, are they:

Family	O	Childminder	O	Nursery	O

3. If other carers, how many hours in total do these other carers look after your child in a typical week?

1-20 hours	O	21-35 hours	O	36+ hours	O

4. Does your child regularly hear a language that is not English?

Yes	O	No	O

5. If yes, for how many hours does your child hear this other language in a typical week:

6. If yes, what is this other language?

7. How many hours per day - including nap time - does your child sleep?

		9	O	10	O	11	O
12	O	13	O	14	O	15 or more	O

You have finished!

You have helped to complete the first large-scale UK-wide study on children's early word learning

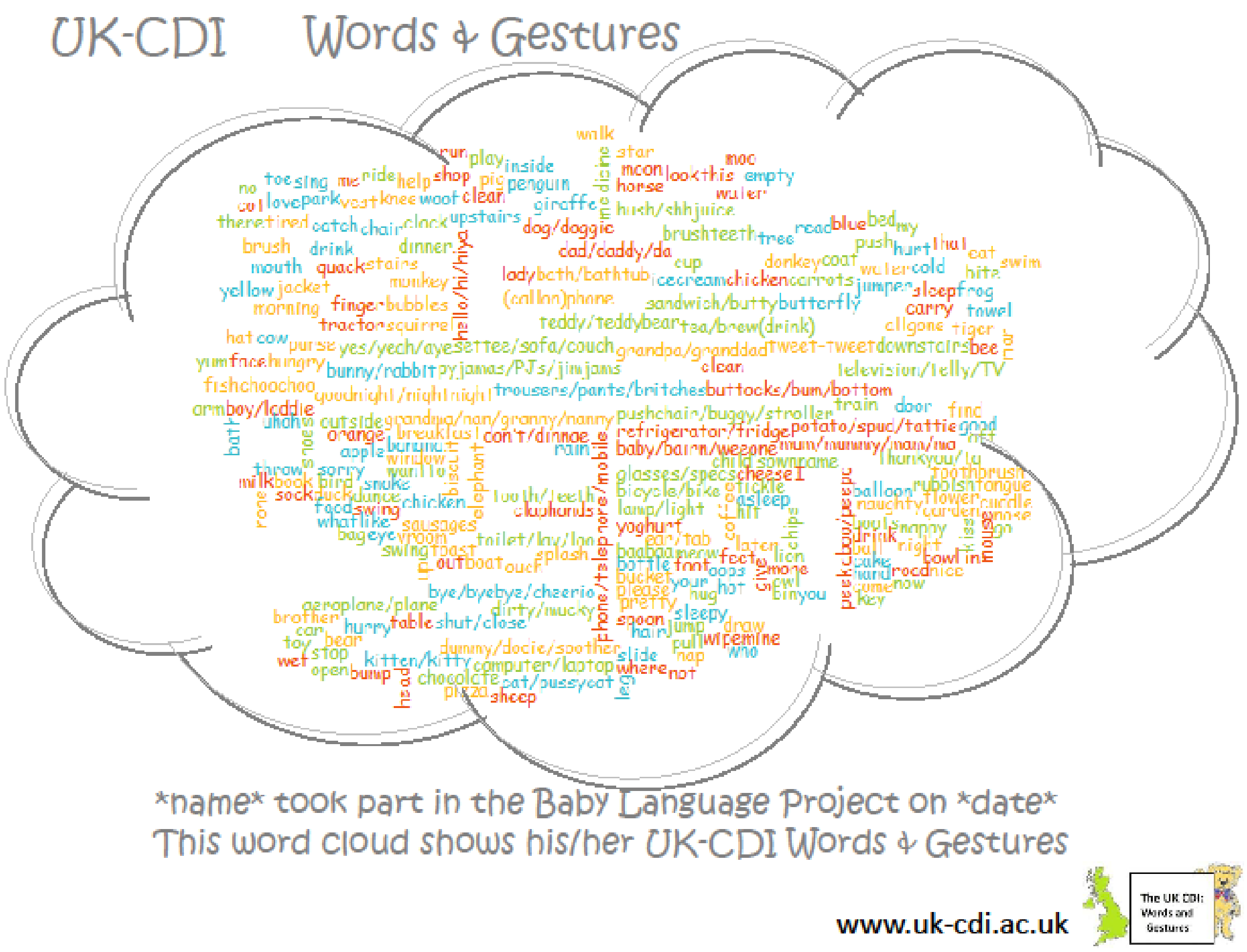
Thank you very much for your time and effort.

Appendix 5 - Details of the adaptation of the Lincoln Toddler CDI (Lincoln TCDI)

Several changes were made with regards to UK spelling, pronunciation and word use (i.e. from airplane to aeroplane / plane; from pajamas to pyjamas, from mommy to mummy). At other times words were replaced in full with the British equivalent (i.e. from rooster to cockerel, from kitty to kitten, firetruck to fire engine, motorcycle to motorbike, stroller to pushchair, sled to sleigh etc.). In many cases the British version offers several words for the same concept, an adaptation first introduced by the Oxford CDIs (Hamilton et al., 2000) in contrast to the American version which often only has one word per concept (e.g. from bunny to bunny/ rabbit, from truck to lorry/truck, from block to block/brick). Other changes involved the replacement of cultural specific words with culturally appropriate words (e.g. in 'Food and Drink', we replaced *applesauce, corn, noodles, pretzels* with *chocolate biscuit, plum, tea and tomato sauce/ketchup*).

The grammar section of the US Toddler CDI has been adopted without any changes, except for the omission of the word 'sitted' in the verb forms of the over-regularisation category. All grammar subsections are phrased in the same way as the original. There are no changes or omissions to any questions. At the end of the questionnaire, additional questions are included in the Lincoln Toddler CDI. The questions ask about the child's first born and sibling status and about childcare provision.

Appendix 6 - Anonymised example of a word-cloud



Appendix 7 - Consent form for video recording

Children's Language Development

Consent Form

PARENT/GUARDIAN **Please read the statements below before signing.**

- I agree that the game session will be video recorded.
- I understand that all data will be kept safe at the Lincoln Infant Lab and will only be accessed by the researchers at the Lincoln Infant Lab.
- I understand that I can withdraw from the study without giving a reason.
- I agree that if I decide to withdraw from the study, I can ask for the data to be destroyed until up to two weeks after participation.

YOUR NAME:	
YOUR RELATIONSHIP TO CHILD:	
SIGNATURE:	X
TODAY'S DATE:	/ /

Appendix 8 - Tests of normality Kolmogorov-Smirnov) for language variables

Tests of Normality

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Phrases Understood 12m	.147	73	.000	.944	73	.003
Gestures 12m	.085	73	.200*	.944	73	.003
Comprehension 12m	.206	73	.000	.784	73	.000
Production 12m	.261	73	.000	.558	73	.000
Phrases Understood 18m	.134	73	.002	.931	73	.001
Gestures 18m	.081	73	.200*	.980	73	.284
Comprehension 18m	.074	73	.200*	.972	73	.104
Production 18m	.184	73	.000	.771	73	.000
PLS Auditory Comprehension 18m	.102	73	.059	.983	73	.425
PLS Expressive Communication 18m	.221	73	.000	.902	73	.000
PLS Total Language 18m	.070	73	.200*	.975	73	.157
Production 24m	.115	73	.019	.958	73	.017
Comprehension 24m	.091	73	.200*	.976	73	.186
Sentence Complexity 24m	.223	73	.000	.802	73	.000
PLS Auditory Comprehension 24m	.105	73	.043	.966	73	.048
PLS Expressive Communication 24m	.091	73	.200*	.953	73	.008
PLS Total Language 24m	.098	73	.079	.966	73	.044
ASQ Communication 36m	.261	73	.000	.806	73	.000
Production 36m	.179	73	.000	.887	73	.000
Sentence Complexity 36m	.185	73	.000	.850	73	.000
Advanced Grammar 36m	.179	73	.000	.864	73	.000

*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction

Appendix 9 - Gender differences in CDI scores

First signs of understanding 12m:

Levene's Test of Equality of Error Variances[a]

Dependent Variable: First signs of understanding 12m

F	df1	df2	Sig.
.323	1	80	.571

Tests the null hypothesis that the error variance of the dependent variable is equal across groups.

a. Design: Intercept + Age_correctedM12 + Gender

Tests of Between-Subjects Effects

Dependent Variable: First signs of understanding 12m

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	.003[a]	2	.002	.069	.933	.002
Intercept	7.141	1	7.141	289.624	.000	.786
Age_correctedM12	.001	1	.001	.060	.808	.001
Gender	.002	1	.002	.077	.782	.001
Error	1.948	79	.025			
Total	728.000	82				
Corrected Total	1.951	81				

a. R Squared = .002 (Adjusted R Squared = -.024)

Gender

Dependent Variable: First signs of understanding 12m

Gender	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
male	2.980[a]	.022	2.935	3.024
female	2.970[a]	.027	2.915	3.024

a. Covariates appearing in the model are evaluated at the following values: preterm children corrected age at 12m = 11.7073.

Phrases understood 12m:

Levene's Test of Equality of Error Variances[a]

Dependent Variable: Phrases understood 12m

F	df1	df2	Sig.
.040	1	80	.842

Tests the null hypothesis that the error variance of the dependent variable is equal across groups.

a. Design: Intercept + Age_correctedM12 + Gender

Tests of Between-Subjects Effects

Dependent Variable: Phrases understood 12m

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	.232[a]	2	.116	1.542	.220	.038
Intercept	.169	1	.169	2.249	.138	.028
Age_correctedM12	.232	1	.232	3.083	.083	.038
Gender	7.781E-5	1	7.781E-5	.001	.974	.000
Error	5.949	79	.075			
Total	84.546	82				
Corrected Total	6.181	81				

a. R Squared = .038 (Adjusted R Squared = .013)

Gender

Dependent Variable: Phrases understood 12m

Gender	Mean	Std. Error	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
male	.977[a]	.039	.899	1.055
female	.979[a]	.048	.884	1.074

a. Covariates appearing in the model are evaluated at the following values: preterm children corrected age at 12m = 11.7073.

Gestures 12m:

Levene's Test of Equality of Error Variances[a]

Dependent Variable: Gestures 12m

F	df1	df2	Sig.
.068	1	80	.795

Tests the null hypothesis that the error variance of the dependent variable is equal across groups.

a. Design: Intercept + Age_correctedM12 + Gender

Tests of Between-Subjects Effects

Dependent Variable: Gestures 12m

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	1299.573[a]	2	649.787	10.965	.000	.217
Intercept	176.152	1	176.152	2.973	.089	.036
Age_correctedM12	1215.817	1	1215.817	20.517	.000	.206
Gender	88.391	1	88.391	1.492	.226	.019
Error	4681.357	79	59.258			
Total	50493.250	82				
Corrected Total	5980.930	81				

a. R Squared = .217 (Adjusted R Squared = .197)

Gender

Dependent Variable: Gestures 12m

Gender	Mean	Std. Error	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
male	22.447[a]	1.100	20.258	24.636
female	24.564[a]	1.340	21.897	27.231

a. Covariates appearing in the model are evaluated at the following values: preterm children corrected age at 12m = 11.7073.

Comprehension 12m:

Levene's Test of Equality of Error Variances[a]

Dependent Variable: Comprehension 12m

F	df1	df2	Sig.
.189	1	80	.665

Tests the null hypothesis that the error variance of the dependent variable is equal across groups.

a. Design: Intercept + Age_correctedM12 + Gender

Tests of Between-Subjects Effects

Dependent Variable: Comprehension 12m

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	.514[a]	2	.257	2.383	.099	.057
Intercept	.730	1	.730	6.770	.011	.079
Age_correctedM12	.501	1	.501	4.648	.034	.056
Gender	.012	1	.012	.108	.744	.001
Error	8.523	79	.108			
Total	249.793	82				
Corrected Total	9.037	81				

a. R Squared = .057 (Adjusted R Squared = .033)

Gender

Dependent Variable: Comprehension 12m

Gender	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
male	1.723[a]	.047	1.630	1.817
female	1.699[a]	.057	1.585	1.813

a. Covariates appearing in the model are evaluated at the following values: preterm children corrected age at 12m = 11.7073.

Production 12m:

Levene's Test of Equality of Error Variances[a]

Dependent Variable: Production 12m

F	df1	df2	Sig.
2.281	1	80	.135

Tests the null hypothesis that the error variance of the dependent variable is equal across groups.

a. Design: Intercept + Age_correctedM12 + Gender

Tests of Between-Subjects Effects

Dependent Variable: Production 12m

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	.970[a]	2	.485	2.289	.108	.055
Intercept	.099	1	.099	.466	.497	.006
Age_correctedM12	.939	1	.939	4.433	.038	.053
Gender	.033	1	.033	.156	.694	.002
Error	16.742	79	.212			
Total	58.969	82				
Corrected Total	17.712	81				

a. R Squared = .055 (Adjusted R Squared = .031)

Gender

Dependent Variable: Production 12m

Gender	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
male	.693[a]	.066	.562	.824
female	.734[a]	.080	.574	.893

a. Covariates appearing in the model are evaluated at the following values: preterm children corrected age at 12m = 11.7073.

Phrases understood 18m:

Levene's Test of Equality of Error Variances[a]

Dependent Variable: Phrases understood 18m

F	df1	df2	Sig.
3.233	1	80	.076

Tests the null hypothesis that the error variance of the dependent variable is equal across groups.

a. Design: Intercept + AllAge_correctedM18 + Gender

Tests of Between-Subjects Effects

Dependent Variable: Phrases understood 18m

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	75.034[a]	2	37.517	1.673	.194	.041
Intercept	93.236	1	93.236	4.159	.045	.050
AllAge_correctedM18	5.747	1	5.747	.256	.614	.003
Gender	66.993	1	66.993	2.988	.088	.036
Error	1771.222	79	22.421			
Total	42197.000	82				
Corrected Total	1846.256	81				

a. R Squared = .041 (Adjusted R Squared = .016)

Gender

Dependent Variable: Phrases understood 18m

Gender	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
male	21.440[a]	.677	20.093	22.787
female	23.286[a]	.825	21.644	24.928

a. Covariates appearing in the model are evaluated at the following values: AllAge_correctedM18 = 17.8293.

Gestures 18m:

Levene's Test of Equality of Error Variances[a]

Dependent Variable: Gestures 18m

F	df1	df2	Sig.
.063	1	78	.803

Tests the null hypothesis that the error variance of the dependent variable is equal across groups.

a. Design: Intercept + AllAge_correctedM18 + Gender

Tests of Between-Subjects Effects

Dependent Variable: Gestures 18m

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	534.922[a]	2	267.461	3.999	.022	.094
Intercept	78.082	1	78.082	1.168	.283	.015
AllAge_correctedM18	25.553	1	25.553	.382	.538	.005
Gender	512.974	1	512.974	7.671	.007	.091
Error	5149.328	77	66.874			
Total	161445.500	80				
Corrected Total	5684.250	79				

a. R Squared = .094 (Adjusted R Squared = .071)

Gender

Dependent Variable: Gestures 18m

Gender	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
male	42.003[a]	1.193	39.628	44.378
female	47.147[a]	1.424	44.313	49.982

a. Covariates appearing in the model are evaluated at the following values: AllAge_correctedM18 = 17.8000.

Comprehension 18m:

Levene's Test of Equality of Error Variances[a]

Dependent Variable: Comprehension 18m

F	df1	df2	Sig.
5.370	1	80	.023

Tests the null hypothesis that the error variance of the dependent variable is equal across groups.

a. Design: Intercept + AllAge_correctedM18 + Gender

Tests of Between-Subjects Effects

Dependent Variable: Comprehension 18m

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	17690.919[a]	2	8845.460	1.200	.307	.029
Intercept	35856.166	1	35856.166	4.866	.030	.058
AllAge_correctedM18	14512.496	1	14512.496	1.969	.164	.024
Gender	2490.681	1	2490.681	.338	.563	.004
Error	582145.679	79	7368.933			
Total	4258581.000	82				
Corrected Total	599836.598	81				

a. R Squared = .029 (Adjusted R Squared = .005)

Gender

Dependent Variable: Comprehension 18m

Gender	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
male	206.702[a]	12.270	182.279	231.126
female	217.957[a]	14.956	188.188	247.726

a. Covariates appearing in the model are evaluated at the following values: AllAge_correctedM18 = 17.8293.

An independent t-test was used for comprehension at 18 months as the Levene's Test of Equality of Error Variances was violated in the ANCOVA:

Group Statistics

	Gender	N	Mean	Std. Deviation	Std. Error Mean
Comprehension 18m	male	49	206.12	96.119	13.731
	female	33	218.82	69.189	12.044

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference Lower	Upper
Comprehension 18m	Equal variances assumed	5.944	.017	-.653	80	.516	-12.696	19.448	-51.398	26.007
	Equal variances not assumed			-.695	79.598	.489	-12.696	18.265	-49.047	23.656

Production 18m:

Levene's Test of Equality of Error Variances[a]

Dependent Variable: Production 18m

F	df1	df2	Sig.
.502	1	80	.481

Tests the null hypothesis that the error variance of the dependent variable is equal across groups.

a. Design: Intercept + AllAge_correctedM18 + Gender

Tests of Between-Subjects Effects

Dependent Variable: Production 18m

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	1.226[a]	2	.613	2.634	.078	.063
Intercept	.072	1	.072	.309	.580	.004
AllAge_correctedM18	.615	1	.615	2.642	.108	.032
Gender	.676	1	.676	2.905	.092	.035
Error	18.382	79	.233			
Total	221.759	82				
Corrected Total	19.608	81				

a. R Squared = .063 (Adjusted R Squared = .039)

Gender

Dependent Variable: Production 18m

Gender	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
male	1.495[a]	.069	1.358	1.633
female	1.681[a]	.084	1.514	1.848

a. Covariates appearing in the model are evaluated at the following values: AllAge_correctedM18 = 17.8293.

Comprehension 24m:

Levene's Test of Equality of Error Variances[a]

Dependent Variable: Comprehension 24m

F	df1	df2	Sig.
4.327	1	80	.041

Tests the null hypothesis that the error variance of the dependent variable is equal across groups.

a. Design: Intercept + Age24 + Gender

Tests of Between-Subjects Effects

Dependent Variable: Comprehension 24m

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	17590.080[a]	2	8795.040	.529	.591	.013
Intercept	15839.644	1	15839.644	.952	.332	.012
Age24	2949.934	1	2949.934	.177	.675	.002
Gender	15149.004	1	15149.004	.911	.343	.011
Error	1314208.371	79	16635.549			
Total	17241337.000	82				
Corrected Total	1331798.451	81				

a. R Squared = .013 (Adjusted R Squared = -.012)

Gender

Dependent Variable: Comprehension 24m

Gender	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
male	429.312[a]	18.432	392.625	465.999
female	457.052[a]	22.463	412.340	501.763

a. Covariates appearing in the model are evaluated at the following values: Age24 = 24.4620.

An independent t-test was used for comprehension at 24 months as the Levene's Test of Equality of Error Variances was violated in the ANCOVA:

Group Statistics

	Gender	N	Mean	Std. Deviation	Std. Error Mean
Comprehension 24m	male	49	429.5102	140.78813	20.11259
	female	33	456.7576	106.90768	18.61024

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Comprehension 24m	Equal variances assumed	4.825	.031	-.943	80	.349	-27.24737	28.89521	-84.75068	30.25594
	Equal variances not assumed			-.994	78.768	.323	-27.24737	27.40177	-81.79173	27.29699

Production 24m:

Levene's Test of Equality of Error Variances[a]

Dependent Variable: Production 24m

F	df1	df2	Sig.
.681	1	80	.412

Tests the null hypothesis that the error variance of the dependent variable is equal across groups.

a. Design: Intercept + Age24 + Gender

Tests of Between-Subjects Effects

Dependent Variable: Production 24m

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	22232.840[a]	2	11116.420	.382	.684	.010
Intercept	2255.749	1	2255.749	.077	.782	.001
Age24	9200.713	1	9200.713	.316	.576	.004
Gender	12146.602	1	12146.602	.417	.520	.005
Error	2301844.538	79	29137.273			
Total	9572073.000	82				
Corrected Total	2324077.378	81				

a. R Squared = .010 (Adjusted R Squared = -.016)

Gender

Dependent Variable: Production 24m

Gender	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
male	287.309[a]	24.393	238.755	335.862
female	312.148[a]	29.729	252.974	371.321

a. Covariates appearing in the model are evaluated at the following values: Age24 = 24.4620.

Production 36m:

Levene's Test of Equality of Error Variances[a]

Dependent Variable: Production 36m

F	df1	df2	Sig.
.560	1	80	.457

Tests the null hypothesis that the error variance of the dependent variable is equal across groups.

a. Design: Intercept + Age_36m + Gender

Tests of Between-Subjects Effects

Dependent Variable: Production 36m

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	.268[a]	2	.134	.409	.666	.010
Intercept	.268	1	.268	.818	.368	.010
Age_36m	.118	1	.118	.360	.550	.005
Gender	.203	1	.203	.620	.433	.008
Error	25.885	79	.328			
Total	147.405	82				
Corrected Total	26.153	81				

a. R Squared = .010 (Adjusted R Squared = -.015)

Gender

Dependent Variable: Production 36m

Gender	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
male	1.258[a]	.083	1.094	1.422
female	1.154[a]	.101	.953	1.355

a. Covariates appearing in the model are evaluated at the following values: Age_36m = 36.7401.

Sentence Complexity 24m:

Levene's Test of Equality of Error Variances[a]

Dependent Variable: Sentence Complexity 24m

F	df1	df2	Sig.
.145	1	80	.705

Tests the null hypothesis that the error variance of the dependent variable is equal across groups.

a. Design: Intercept + Age24 + Gender

Tests of Between-Subjects Effects

Dependent Variable: Sentence Complexity 24m

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	.663[a]	2	.331	1.133	.327	.028
Intercept	.001	1	.001	.004	.950	.000
Age24	.007	1	.007	.023	.881	.000
Gender	.650	1	.650	2.221	.140	.027
Error	23.106	79	.292			
Total	63.566	82				
Corrected Total	23.768	81				

a. R Squared = .028 (Adjusted R Squared = .003)

Gender

Dependent Variable: Sentence Complexity 24m

Gender	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
male	.624[a]	.077	.470	.777
female	.805[a]	.094	.618	.993

a. Covariates appearing in the model are evaluated at the following values: Age24 = 24.4620.

Sentence Complexity 36m:

Levene's Test of Equality of Error Variances[a]

Dependent Variable: Sentence Complexity 36m

F	df1	df2	Sig.
11.556	1	80	.001

Tests the null hypothesis that the error variance of the dependent variable is equal across groups.

a. Design: Intercept + Age_36m + Gender

Tests of Between-Subjects Effects

Dependent Variable: Sentence Complexity 36m

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	.001[a]	2	.000	1.343	.267	.033
Intercept	.080	1	.080	238.336	.000	.751
Age_36m	5.221E-6	1	5.221E-6	.016	.901	.000
Gender	.001	1	.001	2.468	.120	.030
Error	.026	79	.000			
Total	317.573	82				

Corrected Total	.027	81				

a. R Squared = .033 (Adjusted R Squared = .008)

Gender

Dependent Variable: Sentence Complexity 36m

Gender	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
male	1.971[a]	.003	1.965	1.976
female	1.964[a]	.003	1.957	1.970

a. Covariates appearing in the model are evaluated at the following values: Age_36m = 36.7401.

Levene's test was significant, therefore an independent t-test was conducted:

Group Statistics

	Gender	N	Mean	Std. Deviation	Std. Error Mean
Sentence Complexity 36m	male	49	1.9706	.02043	.00292
	female	33	1.9639	.01407	.00245

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
Sentence Complexity 36m	Equal variances assumed	11.283	.001	1.644	80	.104	.00672	.00409	-.00141	.01486
	Equal variances not assumed			1.764	79.942	.081	.00672	.00381	-.00086	.01431

Advanced grammar 36m:

Levene's Test of Equality of Error Variances[a]

Dependent Variable: Advanced grammar 36m

F	df1	df2	Sig.
7.269	1	80	.009

Tests the null hypothesis that the error variance of the dependent variable is equal across groups.

a. Design: Intercept + Age_36m + Gender

Tests of Between-Subjects Effects

Dependent Variable: Advanced grammar 36m

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	.001[a]	2	.000	1.990	.143	.048
Intercept	.072	1	.072	362.689	.000	.821
Age_36m	.000	1	.000	1.262	.265	.016
Gender	.000	1	.000	1.890	.173	.023
Error	.016	79	.000			
Total	316.893	82				
Corrected Total	.016	81				

a. R Squared = .048 (Adjusted R Squared = .024)

Gender

Dependent Variable: Advanced grammar 36m

Gender	Mean	Std. Error	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
male	1.968[a]	.002	1.964	1.972
female	1.963[a]	.002	1.958	1.968

a. Covariates appearing in the model are evaluated at the following values: Age_36m = 36.7401.

Levene's test was significant, therefore an independent t-test was conducted:

Group Statistics

	Gender	N	Mean	Std. Deviation	Std. Error Mean
Advanced grammar 36m	male	49	1.9679	.01619	.00231
	female	33	1.9627	.01022	.00178

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Advanced grammar 36m	Equal variances assumed	7.767	.007	1.646	80	.104	.00523	.00318	-.00109	.01155
	Equal variances not assumed			1.792	79.737	.077	.00523	.00292	-.00058	.01104

Appendix 10 - Gender differences for PLS-5UK scores

Auditory comprehension 18m:

Levene's Test of Equality of Error Variances[a]

Dependent Variable: Auditory comprehension 18m

F	df1	df2	Sig.
2.977	1	77	.088

Tests the null hypothesis that the error variance of the dependent variable is equal across groups.

a. Design: Intercept + AllAge_correctedM18 + Gender

Tests of Between-Subjects Effects

Dependent Variable: Auditory comprehension 18m

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	1303.883[a]	2	651.942	3.135	.049	.076
Intercept	3428.173	1	3428.173	16.483	.000	.178
AllAge_correctedM18	638.055	1	638.055	3.068	.084	.039
Gender	532.153	1	532.153	2.559	.114	.033
Error	15806.193	76	207.976			
Total	898828.000	79				
Corrected Total	17110.076	78				

a. R Squared = .076 (Adjusted R Squared = .052)

Gender

Dependent Variable: Auditory comprehension 18m

Gender	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
male	103.549[a]	2.086	99.394	107.703
female	108.892[a]	2.599	103.717	114.068

a. Covariates appearing in the model are evaluated at the following values: AllAge_correctedM18 = 17.8228.

Expressive communication 18m:

Levene's Test of Equality of Error Variances[a]

Dependent Variable: Expressive communication 18m

F	df1	df2	Sig.
1.378	1	77	.244

Tests the null hypothesis that the error variance of the dependent variable is equal across groups.

a. Design: Intercept + AllAge_correctedM18 + Gender

Tests of Between-Subjects Effects

Dependent Variable: Expressive communication 18m

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	.003[a]	2	.001	1.322	.273	.034
Intercept	.440	1	.440	414.577	.000	.845
AllAge_correctedM18	.001	1	.001	1.103	.297	.014
Gender	.001	1	.001	1.269	.264	.016
Error	.081	76	.001			

Total	318.089	79				
Corrected Total	.083	78				

a. R Squared = .034 (Adjusted R Squared = .008)

Gender

Dependent Variable: Expressive communication 18m

Gender	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
male	2.003[a]	.005	1.994	2.012
female	2.012[a]	.006	2.000	2.023

a. Covariates appearing in the model are evaluated at the following values: AllAge_correctedM18 = 17.8228.

Total Language 18m:

Levene's Test of Equality of Error Variances[a]

Dependent Variable: Total language 18m

F	df1	df2	Sig.
.641	1	77	.426

Tests the null hypothesis that the error variance of the dependent variable is equal across groups.

a. Design: Intercept + AllAge_correctedM18 + Gender

Tests of Between-Subjects Effects

Dependent Variable: Total language 18m

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	761.551[a]	2	380.775	3.197	.046	.078
Intercept	2653.428	1	2653.428	22.280	.000	.227
AllAge_correctedM18	352.065	1	352.065	2.956	.090	.037
Gender	331.179	1	331.179	2.781	.100	.035
Error	9051.335	76	119.097			
Total	864901.000	79				
Corrected Total	9812.886	78				

a. R Squared = .078 (Adjusted R Squared = .053)

Gender

Dependent Variable: Total language 18m

Gender	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
male	102.384[a]	1.578	99.240	105.528
female	106.599[a]	1.966	102.683	110.516

a. Covariates appearing in the model are evaluated at the following values: AllAge_correctedM18 = 17.8228.

PLS 24m:

Group Statistics

	Gender	N	Mean	Std. Deviation	Std. Error Mean
Auditory Comprehension 24m	male	47	100.6170	12.40363	1.80926
	female	31	107.1613	9.81868	1.76349
Expressive Communication 24m	male	47	99.6170	8.86751	1.29346
	female	31	103.7742	10.39137	1.86635
Total Language 24m	male	47	100.1702	10.38879	1.51536
	female	31	105.9677	9.59334	1.72302

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Auditory Comprehension 24m	Equal variances assumed	2.114	.150	-2.470	76	.016	-6.54427	2.64998	-11.82217	-1.26637
	Equal variances not assumed			-2.590	73.375	.012	-6.54427	2.52652	-11.57918	-1.50936
Expressive Communication 24m	Equal variances assumed	.003	.958	-1.892	76	.062	-4.15717	2.19767	-8.53421	.21987
	Equal variances not assumed			-1.831	57.142	.072	-4.15717	2.27075	-8.70402	.38967
Total Language 24m	Equal variances assumed	1.796	.184	-2.485	76	.015	-5.79753	2.33280	-10.44370	-1.15136
	Equal variances not assumed			-2.527	67.874	.014	-5.79753	2.29458	-10.37645	-1.21861

Appendix 11 - Gender differences in terms of ASQ communication scores at 36 months

Group Statistics

	Gender	N	Mean	Std. Deviation	Std. Error Mean
ASQ communication 36m	male	49	1.6839	.07547	.01078
	female	33	1.6656	.05960	.01037

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
ASQ communication 36m	Equal variances assumed	3.054	.084	1.166	80	.247	.01827	.01566	-.01290	.04944
	Equal variances not assumed			1.221	77.883	.226	.01827	.01496	-.01152	.04806

Appendix 12 - The effect of gender and maternal education on language ability groups

The effect of gender on language ability groups

Comprehension levels at 12m

Gender * Comprehension levels at 12m Crosstabulation

			Comprehension levels at 12m				
			1st - 25th percentile	26th - 49th percentile	50th - 74th percentile	75th - 99th percentile	Total
Gender	male	Count	10	15	8	6	39
		% within Gender	25.6%	38.5%	20.5%	15.4%	100.0%

		% within Comprehension levels at 12m	58.8%	65.2%	47.1%	60.0%	58.2%
		% of Total	14.9%	22.4%	11.9%	9.0%	58.2%
	female	Count	7	8	9	4	28
		% within Gender	25.0%	28.6%	32.1%	14.3%	100.0%
		% within Comprehension levels at 12m	41.2%	34.8%	52.9%	40.0%	41.8%
		% of Total	10.4%	11.9%	13.4%	6.0%	41.8%
Total		Count	17	23	17	10	67
		% within Gender	25.4%	34.3%	25.4%	14.9%	100.0%
		% within Comprehension levels at 12m	100.0%	100.0%	100.0%	100.0%	100.0%
		% of Total	25.4%	34.3%	25.4%	14.9%	100.0%

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	1.349[a]	3	.718
Likelihood Ratio	1.344	3	.719
Linear-by-Linear Association	.160	1	.689
N of Valid Cases	67		

a. 1 cells (12.5%) have expected count less than 5. The minimum expected count is 4.18.

Symmetric Measures

		Value	Approx. Sig.
Nominal by Nominal	Phi	.142	.718
	Cramer's V	.142	.718
N of Valid Cases		67	

Production levels at 12m

Gender * production levels at 12mCrosstabulation

			production levels at 12m				Total
			1st - 25th percentile	26th - 49th percentile	50th - 74th percentile	75th - 99th percentile	
Gender	male	Count	13	5	11	10	39
		% within Gender	33.3%	12.8%	28.2%	25.6%	100.0%
		% within production levels at 12m	76.5%	33.3%	55.0%	66.7%	58.2%
		% of Total	19.4%	7.5%	16.4%	14.9%	58.2%
	female	Count	4	10	9	5	28
		% within Gender	14.3%	35.7%	32.1%	17.9%	100.0%
		% within production levels at 12m	23.5%	66.7%	45.0%	33.3%	41.8%
		% of Total	6.0%	14.9%	13.4%	7.5%	41.8%
Total		Count	17	15	20	15	67
		% within Gender	25.4%	22.4%	29.9%	22.4%	100.0%
		% within production levels at 12m	100.0%	100.0%	100.0%	100.0%	100.0%
		% of Total	25.4%	22.4%	29.9%	22.4%	100.0%

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	6.672[a]	3	.083
Likelihood Ratio	6.801	3	.079
Linear-by-Linear Association	.073	1	.787
N of Valid Cases	67		

a. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 6.27.

Symmetric Measures

		Value	Approx. Sig.
Nominal by Nominal	Phi	.316	.083
	Cramer's V	.316	.083
N of Valid Cases		67	

Gesture levels at 12m

Gender * Gesture levels at 12m Crosstabulation

			Gesture levels at 12m				
			1st - 25th percentile	26th - 49th percentile	50th - 74th percentile	75th - 99th percentile	Total
Gender	male	Count	8	9	10	12	39
		% within Gender	20.5%	23.1%	25.6%	30.8%	100.0%
		% within Gesture levels at 12m	72.7%	60.0%	58.8%	50.0%	58.2%
		% of Total	11.9%	13.4%	14.9%	17.9%	58.2%
	female	Count	3	6	7	12	28
		% within Gender	10.7%	21.4%	25.0%	42.9%	100.0%
		% within Gesture levels at 12m	27.3%	40.0%	41.2%	50.0%	41.8%
		% of Total	4.5%	9.0%	10.4%	17.9%	41.8%
Total		Count	11	15	17	24	67
		% within Gender	16.4%	22.4%	25.4%	35.8%	100.0%
		% within Gesture levels at 12m	100.0%	100.0%	100.0%	100.0%	100.0%
		% of Total	16.4%	22.4%	25.4%	35.8%	100.0%

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	1.640[a]	3	.650
Likelihood Ratio	1.680	3	.641
Linear-by-Linear Association	1.485	1	.223
N of Valid Cases	67		

a. 1 cells (12.5%) have expected count less than 5. The minimum expected count is 4.60.

Symmetric Measures

		Value	Approx. Sig.
Nominal by Nominal	Phi	.156	.650
	Cramer's V	.156	.650
N of Valid Cases		67	

Comprehension levels at 18m

Gender * comprehension levels at 18mCrosstabulation

			comprehension levels at 18m				Total
			1st - 25th percentile	26th - 49th percentile	50th - 74th percentile	75th - 99th percentile	
Gender	male	Count	16	7	7	9	39
		% within Gender	41.0%	17.9%	17.9%	23.1%	100.0%
		% within comprehension levels at 18m	72.7%	43.8%	46.7%	64.3%	58.2%
		% of Total	23.9%	10.4%	10.4%	13.4%	58.2%
	female	Count	6	9	8	5	28
		% within Gender	21.4%	32.1%	28.6%	17.9%	100.0%
		% within comprehension levels at 18m	27.3%	56.3%	53.3%	35.7%	41.8%
		% of Total	9.0%	13.4%	11.9%	7.5%	41.8%
Total		Count	22	16	15	14	67
		% within Gender	32.8%	23.9%	22.4%	20.9%	100.0%
		% within comprehension levels at 18m	100.0%	100.0%	100.0%	100.0%	100.0%
		% of Total	32.8%	23.9%	22.4%	20.9%	100.0%

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	4.315[a]	3	.229
Likelihood Ratio	4.379	3	.223
Linear-by-Linear Association	.487	1	.485
N of Valid Cases	67		

a. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 5.85.

Symmetric Measures

		Value	Approx. Sig.
Nominal by Nominal	Phi	.254	.229
	Cramer's V	.254	.229
N of Valid Cases		67	

Production levels at 18m

Gender * production levels at 18m Crosstabulation

			production levels at 18m				Total
			1st - 25th percentile	26th - 49th percentile	50th - 74th percentile	75th - 99th percentile	
Gender	male	Count	11	13	9	6	39
		% within Gender	28.2%	33.3%	23.1%	15.4%	100.0%
		% within production levels at 18m	73.3%	61.9%	56.3%	40.0%	58.2%
		% of Total	16.4%	19.4%	13.4%	9.0%	58.2%
	female	Count	4	8	7	9	28
		% within Gender	14.3%	28.6%	25.0%	32.1%	100.0%
		% within production levels at 18m	26.7%	38.1%	43.8%	60.0%	41.8%
		% of Total	6.0%	11.9%	10.4%	13.4%	41.8%
Total		Count	15	21	16	15	67
		% within Gender	22.4%	31.3%	23.9%	22.4%	100.0%
		% within production levels at 18m	100.0%	100.0%	100.0%	100.0%	100.0%
		% of Total	22.4%	31.3%	23.9%	22.4%	100.0%

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	3.598[a]	3	.308
Likelihood Ratio	3.639	3	.303
Linear-by-Linear Association	3.419	1	.064
N of Valid Cases	67		

a. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 6.27.

Symmetric Measures

		Value	Approx. Sig.
Nominal by Nominal	Phi	.232	.308
	Cramer's V	.232	.308
N of Valid Cases		67	

Gesture levels at 18m

Gender * gesture levels at 18mCrosstabulation

			gesture levels at 18m				
			1st - 25th percentile	26th - 49th percentile	50th - 74th percentile	75th - 99th percentile	Total
Gender	male	Count	10	12	10	7	39
		% within Gender	25.6%	30.8%	25.6%	17.9%	100.0%
		% within gesture levels at 18m	83.3%	66.7%	52.6%	38.9%	58.2%
		% of Total	14.9%	17.9%	14.9%	10.4%	58.2%
	female	Count	2	6	9	11	28
		% within Gender	7.1%	21.4%	32.1%	39.3%	100.0%
		% within gesture levels at 18m	16.7%	33.3%	47.4%	61.1%	41.8%
		% of Total	3.0%	9.0%	13.4%	16.4%	41.8%
Total		Count	12	18	19	18	67
		% within Gender	17.9%	26.9%	28.4%	26.9%	100.0%
		% within gesture levels at 18m	100.0%	100.0%	100.0%	100.0%	100.0%
		% of Total	17.9%	26.9%	28.4%	26.9%	100.0%

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	6.648[a]	3	.084
Likelihood Ratio	6.996	3	.072
Linear-by-Linear Association	6.534	1	.011
N of Valid Cases	67		

a. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 5.01.

Symmetric Measures

		Value	Approx. Sig.
Nominal by Nominal	Phi	.315	.084
	Cramer's V	.315	.084
N of Valid Cases		67	

The effect of maternal education on language ability groups

Comprehension levels at 12m

Maternal education* Comprehension levels at 12m Crosstabulation

			Comprehension levels at 12m				Total
			1st - 25th percentile	26th - 49th percentile	50th - 74th percentile	75th - 99th percentile	
Maternal education	Low level	Count	4	9	9	2	24
		% within Maternal education	16.7%	37.5%	37.5%	8.3%	100.0%
		% within Comprehension levels at 12m	23.5%	39.1%	52.9%	20.0%	35.8%
		% of Total	6.0%	13.4%	13.4%	3.0%	35.8%
	High level	Count	13	14	8	8	43
		% within Maternal education	30.2%	32.6%	18.6%	18.6%	100.0%
		% within Comprehension levels at 12m	76.5%	60.9%	47.1%	80.0%	64.2%
		% of Total	19.4%	20.9%	11.9%	11.9%	64.2%
Total		Count	17	23	17	10	67
		% within Maternal education	25.4%	34.3%	25.4%	14.9%	100.0%
		% within Comprehension levels at 12m	100.0%	100.0%	100.0%	100.0%	100.0%
		% of Total	25.4%	34.3%	25.4%	14.9%	100.0%

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	4.483[a]	3	.214
Likelihood Ratio	4.564	3	.207
Linear-by-Linear Association	.212	1	.645
N of Valid Cases	67		

a. 1 cells (12.5%) have expected count less than 5. The minimum expected count is 3.58.

Symmetric Measures

		Value	Approx. Sig.
Nominal by Nominal	Phi	.259	.214
	Cramer's V	.259	.214
N of Valid Cases		67	

Production levels at 12m

Maternal education * production levels at 12mCrosstabulation

			production levels at 12m				Total
			1st - 25th percentile	26th - 49th percentile	50th - 74th percentile	75th - 99th percentile	
Maternal education	Low level	Count	6	7	7	4	24
		% within Maternal education	25.0%	29.2%	29.2%	16.7%	100.0%
		% within production levels at 12m	35.3%	46.7%	35.0%	26.7%	35.8%
		% of Total	9.0%	10.4%	10.4%	6.0%	35.8%
	High level	Count	11	8	13	11	43
		% within Maternal education	25.6%	18.6%	30.2%	25.6%	100.0%
		% within production levels at 12m	64.7%	53.3%	65.0%	73.3%	64.2%
		% of Total	16.4%	11.9%	19.4%	16.4%	64.2%
Total		Count	17	15	20	15	67
		% within Maternal education	25.4%	22.4%	29.9%	22.4%	100.0%
		% within production levels at 12m	100.0%	100.0%	100.0%	100.0%	100.0%
		% of Total	25.4%	22.4%	29.9%	22.4%	100.0%

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	1.322[a]	3	.724
Likelihood Ratio	1.322	3	.724
Linear-by-Linear Association	.422	1	.516
N of Valid Cases	67		

a. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 5.37.

Symmetric Measures

		Value	Approx. Sig.
Nominal by Nominal	Phi	.140	.724
	Cramer's V	.140	.724
N of Valid Cases		67	

Gesture levels at 12m

Maternal education * Gesture levels at 12m Crosstabulation

			Gesture levels at 12m				
			1st - 25th percentile	26th - 49th percentile	50th - 74th percentile	75th - 99th percentile	Total
Maternal education	Low level	Count	5	3	6	10	24
		% within Maternal education	20.8%	12.5%	25.0%	41.7%	100.0%
		% within Gesture levels at 12m	45.5%	20.0%	35.3%	41.7%	35.8%
		% of Total	7.5%	4.5%	9.0%	14.9%	35.8%
	High level	Count	6	12	11	14	43
		% within Maternal education	14.0%	27.9%	25.6%	32.6%	100.0%
		% within Gesture levels at 12m	54.5%	80.0%	64.7%	58.3%	64.2%
		% of Total	9.0%	17.9%	16.4%	20.9%	64.2%
Total		Count	11	15	17	24	67
		% within Maternal education	16.4%	22.4%	25.4%	35.8%	100.0%
		% within Gesture levels at 12m	100.0%	100.0%	100.0%	100.0%	100.0%
		% of Total	16.4%	22.4%	25.4%	35.8%	100.0%

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	2.436[a]	3	.487
Likelihood Ratio	2.573	3	.462
Linear-by-Linear Association	.146	1	.702
N of Valid Cases	67		

a. 1 cells (12.5%) have expected count less than 5. The minimum expected count is 3.94.

Symmetric Measures

		Value	Approx. Sig.
Nominal by Nominal	Phi	.191	.487
	Cramer's V	.191	.487
N of Valid Cases		67	

Comprehension levels at 18m

Maternal education * comprehension levels at 18mCrosstabulation

			comprehension levels at 18m				Total
			1st - 25th percentile	26th - 49th percentile	50th - 74th percentile	75th - 99th percentile	
Maternal education	Low level	Count	8	7	3	6	24
		% within Maternal education	33.3%	29.2%	12.5%	25.0%	100.0%
		% within comprehension levels at 18m	36.4%	43.8%	20.0%	42.9%	35.8%
		% of Total	11.9%	10.4%	4.5%	9.0%	35.8%
	High level	Count	14	9	12	8	43
		% within Maternal education	32.6%	20.9%	27.9%	18.6%	100.0%
		% within comprehension levels at 18m	63.6%	56.3%	80.0%	57.1%	64.2%
		% of Total	20.9%	13.4%	17.9%	11.9%	64.2%
Total		Count	22	16	15	14	67
		% within Maternal education	32.8%	23.9%	22.4%	20.9%	100.0%
		% within comprehension levels at 18m	100.0%	100.0%	100.0%	100.0%	100.0%
		% of Total	32.8%	23.9%	22.4%	20.9%	100.0%

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	2.375[a]	3	.498
Likelihood Ratio	2.514	3	.473
Linear-by-Linear Association	.014	1	.907
N of Valid Cases	67		

a. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 5.01.

Symmetric Measures

		Value	Approx. Sig.
Nominal by Nominal	Phi	.188	.498
	Cramer's V	.188	.498
N of Valid Cases		67	

Production levels at 18m

Maternal education * production levels at 18m Crosstabulation

			production levels at 18m				
			1st - 25th percentile	26th - 49th percentile	50th - 74th percentile	75th - 99th percentile	Total
Maternal education	Low level	Count	5	8	6	5	24
		% within Maternal education	20.8%	33.3%	25.0%	20.8%	100.0%
		% within production levels at 18m	33.3%	38.1%	37.5%	33.3%	35.8%
		% of Total	7.5%	11.9%	9.0%	7.5%	35.8%
	High level	Count	10	13	10	10	43
		% within Maternal education	23.3%	30.2%	23.3%	23.3%	100.0%
		% within production levels at 18m	66.7%	61.9%	62.5%	66.7%	64.2%
		% of Total	14.9%	19.4%	14.9%	14.9%	64.2%
Total		Count	15	21	16	15	67
		% within Maternal education	22.4%	31.3%	23.9%	22.4%	100.0%
		% within production levels at 18m	100.0%	100.0%	100.0%	100.0%	100.0%
		% of Total	22.4%	31.3%	23.9%	22.4%	100.0%

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	.148[a]	3	.986
Likelihood Ratio	.148	3	.986
Linear-by-Linear Association	.001	1	.980
N of Valid Cases	67		

a. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 5.37.

Symmetric Measures

		Value	Approx. Sig.
Nominal by Nominal	Phi	.047	.986
	Cramer's V	.047	.986
N of Valid Cases		67	

Gesture levels at 18m

Maternal education * gesture levels at 18mCrosstabulation

			gesture levels at 18m				
			1st - 25th percentile	26th - 49th percentile	50th - 74th percentile	75th - 99th percentile	Total
Maternal education	Low level	Count	4	7	3	10	24
		% within Maternal education	16.7%	29.2%	12.5%	41.7%	100.0%
		% within gesture levels at 18m	33.3%	38.9%	15.8%	55.6%	35.8%
		% of Total	6.0%	10.4%	4.5%	14.9%	35.8%
	High level	Count	8	11	16	8	43
		% within Maternal education	18.6%	25.6%	37.2%	18.6%	100.0%
		% within gesture levels at 18m	66.7%	61.1%	84.2%	44.4%	64.2%
		% of Total	11.9%	16.4%	23.9%	11.9%	64.2%
Total		Count	12	18	19	18	67
		% within Maternal education	17.9%	26.9%	28.4%	26.9%	100.0%
		% within gesture levels at 18m	100.0%	100.0%	100.0%	100.0%	100.0%
		% of Total	17.9%	26.9%	28.4%	26.9%	100.0%

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	6.472[a]	3	.091
Likelihood Ratio	6.781	3	.079
Linear-by-Linear Association	.735	1	.391
N of Valid Cases	67		

a. 1 cells (12.5%) have expected count less than 5. The minimum expected count is 4.30.

Symmetric Measures

		Value	Approx. Sig.
Nominal by Nominal	Phi	.311	.091
	Cramer's V	.311	.091
N of Valid Cases		67	

Appendix 13 - Effect of ability group on communication variables between 12 – 36 months

MANOVA

Production levels at 12m

Between-Subjects Factors

		Value Label	N
production levels at 12m	1.00	1st - 25th percentile	17
	2.00	26th - 49th percentile	15
	3.00	50th - 74th percentile	20
	4.00	75th - 99th percentile	15

Descriptive Statistics

	production levels at 12m	Mean	Std. Deviation	N
Zscore comprehension 12m	1st - 25th percentile	-.4294	.96871	17
	26th - 49th percentile	-.1587	.78896	15
	50th - 74th percentile	-.1267	.81737	20
	75th - 99th percentile	1.0623	1.05681	15
	Total	.0555	1.04889	67

Zscore gestures 12m	1st - 25th percentile	-.2333	.98649	17
	26th - 49th percentile	.0104	.92632	15
	50th - 74th percentile	-.1359	.77788	20
	75th - 99th percentile	.4351	1.27569	15
	Total	.0000	1.00000	67
Zscore comprehension 18m	1st - 25th percentile	-.4499	.90355	17
	26th - 49th percentile	-.3158	.67168	15
	50th - 74th percentile	-.0842	.77428	20
	75th - 99th percentile	.9379	1.10055	15
	Total	.0000	1.00000	67
Zscore production 18m	1st - 25th percentile	-.6223	.89167	17
	26th - 49th percentile	-.1530	.83327	15
	50th - 74th percentile	.1625	.68120	20
	75th - 99th percentile	.8942	1.24745	15
	Total	.0566	1.04654	67
Zscore gestures 18m	1st - 25th percentile	-.1487	1.07669	17
	26th - 49th percentile	.2216	.81079	15
	50th - 74th percentile	-.2118	.85959	20
	75th - 99th percentile	.2293	1.23667	15
	Total	.0000	1.00000	67
Zscore Auditory Comprehension 18m	1st - 25th percentile	-.2263	1.15651	17
	26th - 49th percentile	.1270	.84461	15
	50th - 74th percentile	-.1580	.92255	20
	75th - 99th percentile	.3402	1.03832	15
	Total	.0000	1.00000	67
Zscore Expressive Communication 18m	1st - 25th percentile	-.4479	1.09068	17
	26th - 49th percentile	.0867	.99870	15
	50th - 74th percentile	.0047	.71788	20
	75th - 99th percentile	.6159	.82129	15
	Total	.0451	.96382	67
Zscore comprehension 24m	1st - 25th percentile	-.5571	.96072	17
	26th - 49th percentile	-.1299	.86212	15
	50th - 74th percentile	-.0113	.90543	20
	75th - 99th percentile	.7763	.87344	15
	Total	.0000	1.00000	67
Zscore production 24m	1st - 25th percentile	-.4425	1.14323	17
	26th - 49th percentile	.2401	1.10465	15
	50th - 74th percentile	-.0858	.92708	20
	75th - 99th percentile	.6238	.77207	15
	Total	.0555	1.04911	67
Zscore sentence Complexity 24m	1st - 25th percentile	-.5265	.61487	17
	26th - 49th percentile	.0949	1.09277	15

	50th - 74th percentile	.0353	.79969	20
	75th - 99th percentile	.8145	1.05910	15
	Total	.0806	.99246	67
Zscore Auditory Comprehension 24m	1st - 25th percentile	-.1061	1.39316	17
	26th - 49th percentile	.2152	.75308	15
	50th - 74th percentile	-.1624	.79116	20
	75th - 99th percentile	.1215	.98118	15
	Total	.0000	1.00000	67
Zscore Expressive Communication 24m	1st - 25th percentile	-.3931	1.29432	17
	26th - 49th percentile	.2851	.76869	15
	50th - 74th percentile	-.1773	.86539	20
	75th - 99th percentile	.3967	.83918	15
	Total	.0000	1.00000	67
Zscore ASQ 36m	1st - 25th percentile	-.1497	1.04541	17
	26th - 49th percentile	-.1831	.70656	15
	50th - 74th percentile	.0270	.82418	20
	75th - 99th percentile	.2927	.86050	15
	Total	-.0054	.86977	67
Zscore production 36m	1st - 25th percentile	-.4025	.98374	17
	26th - 49th percentile	.2669	.89274	15
	50th - 74th percentile	-.1697	.82728	20
	75th - 99th percentile	.5246	.95727	15
	Total	.0244	.96107	67
Zscore sentence Complexity 36m	1st - 25th percentile	-.6867	.92112	17
	26th - 49th percentile	.2154	.62343	15
	50th - 74th percentile	.0785	.87864	20
	75th - 99th percentile	.4688	.68466	15
	Total	.0023	.89086	67

Box's Test of Equality of Covariance Matrices[a]

Box's M	270.840
F	1.176
df1	120
df2	3583.567
Sig.	.095

Tests the null hypothesis that the observed covariance matrices of the dependent variables are equal across groups.

a. Design: Intercept + production levels at 12m

Multivariate Tests[a]

Effect		Value	F	Hypothesis df	Error df	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power[d]
Intercept	Pillai's Trace	.058	.200[b]	15.000	49.000	.999	.058	2.994	.118
	Wilks' Lambda	.942	.200[b]	15.000	49.000	.999	.058	2.994	.118
	Hotelling's Trace	.061	.200[b]	15.000	49.000	.999	.058	2.994	.118
	Roy's Largest Root	.061	.200[b]	15.000	49.000	.999	.058	2.994	.118
production levels at 12m	Pillai's Trace	1.004	1.710	45.000	153.000	.009	.335	76.956	.998
	Wilks' Lambda	.259	1.869	45.000	146.347	.003	.362	83.101	.999
	Hotelling's Trace	1.929	2.043	45.000	143.000	.001	.391	91.925	1.000
	Roy's Largest Root	1.364	4.639[c]	15.000	51.000	.000	.577	69.588	1.000

a. Design: Intercept + production levels at 12m

b. Exact statistic

c. The statistic is an upper bound on F that yields a lower bound on the significance level.

d. Computed using alpha = .05

Levene's Test of Equality of Error Variances[a]

	F	df1	df2	Sig.
Zscore comprehension 12m	.636	3	63	.594
Zscore gestures 12m	1.049	3	63	.377
Zscore comprehension 18m	1.794	3	63	.157
Zscore production 18m	2.194	3	63	.098
Zscore gestures 18m	.683	3	63	.566

Zscore Auditory Comprehension 18m	.492	3	63	.689
Zscore Expressive Communication 18m	.979	3	63	.408
Zscore comprehension 24m	.043	3	63	.988
Zscore production 24m	.571	3	63	.636
Zscore sentence Complexity 24m	1.758	3	63	.164
Zscore Auditory Comprehension 24m	.878	3	63	.458
Zscore Expressive Communication 24m	1.152	3	63	.335
Zscore ASQ 36m	1.305	3	63	.281
Zscore production 36m	.481	3	63	.697
Zscore sentence Complexity 36m	1.363	3	63	.262

Tests the null hypothesis that the error variance of the dependent variable is equal across groups.

a. Design: Intercept + production levels at 12m

Tests of Between-Subjects Effects

Source	Dependent Variable	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power[p]
Corrected Model	Zscore comprehension 12m	20.553[a]	3	6.851	8.291	.000	.283	24.873	.990
	Zscore gestures 12m	4.136[b]	3	1.379	1.404	.250	.063	4.212	.355
	Zscore comprehension 18m	18.274[c]	3	6.091	8.041	.000	.277	24.122	.987
	Zscore production 18m	19.242[d]	3	6.414	7.618	.000	.266	22.854	.983
	Zscore gestures 18m	2.798[e]	3	.933	.930	.432	.042	2.789	.243
	Zscore Auditory Comprehension 18m	3.348[f]	3	1.116	1.122	.347	.051	3.367	.289
	Zscore Expressive Communication 18m	9.079[g]	3	3.026	3.650	.017	.148	10.950	.774
	Zscore comprehension 24m	14.570[h]	3	4.857	5.949	.001	.221	17.848	.945
	Zscore production 24m	9.971[i]	3	3.324	3.341	.025	.137	10.023	.733
	Zscore sentence Complexity 24m	14.387[j]	3	4.796	5.968	.001	.221	17.905	.946
	Zscore Auditory Comprehension 24m	1.635[k]	3	.545	.533	.661	.025	1.600	.153
	Zscore Expressive Communication 24m	6.835[l]	3	2.278	2.426	.074	.104	7.278	.579

	Zscore ASQ 36m	2.181^{m}	3	.727	.959	.418	.044	2.878	.250
	Zscore production 36m	8.486^{n}	3	2.829	3.396	.023	.139	10.189	.741
	Zscore sentence Complexity 36m	12.132^{o}	3	4.044	6.330	.001	.232	18.990	.957
Intercept	Zscore comprehension 12m	.499	1	.499	.604	.440	.009	.604	.119
	Zscore gestures 12m	.024	1	.024	.025	.876	.000	.025	.053
	Zscore comprehension 18m	.032	1	.032	.042	.838	.001	.042	.055
	Zscore production 18m	.327	1	.327	.389	.535	.006	.389	.094
	Zscore gestures 18m	.034	1	.034	.034	.855	.001	.034	.054
	Zscore Auditory Comprehension 18m	.028	1	.028	.029	.866	.000	.029	.053
	Zscore Expressive Communication 18m	.278	1	.278	.335	.565	.005	.335	.088
	Zscore comprehension 24m	.025	1	.025	.031	.861	.000	.031	.053
	Zscore production 24m	.465	1	.465	.468	.497	.007	.468	.103
	Zscore sentence Complexity 24m	.722	1	.722	.899	.347	.014	.899	.154
	Zscore Auditory Comprehension 24m	.019	1	.019	.019	.891	.000	.019	.052
	Zscore Expressive Communication 24m	.051	1	.051	.055	.816	.001	.055	.056
	Zscore ASQ 36m	.001	1	.001	.001	.976	.000	.001	.050
	Zscore production 36m	.199	1	.199	.239	.627	.004	.239	.077
	Zscore sentence Complexity 36m	.024	1	.024	.037	.848	.001	.037	.054
production levels at 12m	Zscore comprehension 12m	20.553	3	6.851	8.291	.000	.283	24.873	.990
	Zscore gestures 12m	4.136	3	1.379	1.404	.250	.063	4.212	.355
	Zscore comprehension 18m	18.274	3	6.091	8.041	.000	.277	24.122	.987
	Zscore production 18m	19.242	3	6.414	7.618	.000	.266	22.854	.983
	Zscore gestures 18m	2.798	3	.933	.930	.432	.042	2.789	.243
	Zscore Auditory Comprehension 18m	3.348	3	1.116	1.122	.347	.051	3.367	.289
	Zscore Expressive Communication 18m	9.079	3	3.026	3.650	.017	.148	10.950	.774
	Zscore comprehension 24m	14.570	3	4.857	5.949	.001	.221	17.848	.945
	Zscore production 24m	9.971	3	3.324	3.341	.025	.137	10.023	.733
	Zscore sentence Complexity 24m	14.387	3	4.796	5.968	.001	.221	17.905	.946
	Zscore Auditory Comprehension 24m	1.635	3	.545	.533	.661	.025	1.600	.153
	Zscore Expressive Communication 24m	6.835	3	2.278	2.426	.074	.104	7.278	.579
	Zscore ASQ 36m	2.181	3	.727	.959	.418	.044	2.878	.250

	Zscore production 36m	8.486	3	2.829	3.396	.023	.139	10.189	.741
	Zscore sentence Complexity 36m	12.132	3	4.044	6.330	.001	.232	18.990	.957
Error	Zscore comprehension 12m	52.058	63	.826					
	Zscore gestures 12m	61.864	63	.982					
	Zscore comprehension 18m	47.726	63	.758					
	Zscore production 18m	53.044	63	.842					
	Zscore gestures 18m	63.202	63	1.003					
	Zscore Auditory Comprehension 18m	62.652	63	.994					
	Zscore Expressive Communication 18m	52.232	63	.829					
	Zscore comprehension 24m	51.430	63	.816					
	Zscore production 24m	62.670	63	.995					
	Zscore sentence Complexity 24m	50.621	63	.804					
	Zscore Auditory Comprehension 24m	64.365	63	1.022					
	Zscore Expressive Communication 24m	59.165	63	.939					
	Zscore ASQ 36m	47.748	63	.758					
	Zscore production 36m	52.474	63	.833					
	Zscore sentence Complexity 36m	40.247	63	.639					
Total	Zscore comprehension 12m	72.818	67						
	Zscore gestures 12m	66.000	67						
	Zscore comprehension 18m	66.000	67						
	Zscore production 18m	72.501	67						
	Zscore gestures 18m	66.000	67						
	Zscore Auditory Comprehension 18m	66.000	67						
	Zscore Expressive Communication 18m	61.446	67						
	Zscore comprehension 24m	66.000	67						
	Zscore production 24m	72.848	67						
	Zscore sentence Complexity 24m	65.443	67						
	Zscore Auditory Comprehension 24m	66.000	67						
	Zscore Expressive Communication 24m	66.000	67						
	Zscore ASQ 36m	49.931	67						
	Zscore production 36m	61.001	67						

	Zscore sentence Complexity 36m	52.380	67						
Corrected Total	Zscore comprehension 12m	72.612	66						
	Zscore gestures 12m	66.000	66						
	Zscore comprehension 18m	66.000	66						
	Zscore production 18m	72.287	66						
	Zscore gestures 18m	66.000	66						
	Zscore Auditory Comprehension 18m	66.000	66						
	Zscore Expressive Communication 18m	61.310	66						
	Zscore comprehension 24m	66.000	66						
	Zscore production 24m	72.641	66						
	Zscore sentence Complexity 24m	65.008	66						
	Zscore Auditory Comprehension 24m	66.000	66						
	Zscore Expressive Communication 24m	66.000	66						
	Zscore ASQ 36m	49.929	66						
	Zscore production 36m	60.961	66						
	Zscore sentence Complexity 36m	52.379	66						

a. R Squared = .283 (Adjusted R Squared = .249)

b. R Squared = .063 (Adjusted R Squared = .018)

c. R Squared = .277 (Adjusted R Squared = .242)

d. R Squared = .266 (Adjusted R Squared = .231)

e. R Squared = .042 (Adjusted R Squared = -.003)

f. R Squared = .051 (Adjusted R Squared = .006)

g. R Squared = .148 (Adjusted R Squared = .108)

h. R Squared = .221 (Adjusted R Squared = .184)

i. R Squared = .137 (Adjusted R Squared = .096)

j. R Squared = .221 (Adjusted R Squared = .184)

k. R Squared = .025 (Adjusted R Squared = -.022)

l. R Squared = .104 (Adjusted R Squared = .061)

m. R Squared = .044 (Adjusted R Squared = -.002)

n. R Squared = .139 (Adjusted R Squared = .098)

o. R Squared = .232 (Adjusted R Squared = .195)

p. Computed using alpha = .05

production levels at 12m

Dependent Variable	production levels at 12m	Mean	Std. Error	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
Zscore comprehension 12m	1st - 25th percentile	-.429	.220	-.870	.011
	26th - 49th percentile	-.159	.235	-.628	.310
	50th - 74th percentile	-.127	.203	-.533	.280
	75th - 99th percentile	1.062	.235	.593	1.531
Zscore gestures 12m	1st - 25th percentile	-.233	.240	-.714	.247
	26th - 49th percentile	.010	.256	-.501	.522
	50th - 74th percentile	-.136	.222	-.579	.307
	75th - 99th percentile	.435	.256	-.076	.946
Zscore comprehension 18m	1st - 25th percentile	-.450	.211	-.872	-.028
	26th - 49th percentile	-.316	.225	-.765	.133
	50th - 74th percentile	-.084	.195	-.473	.305
	75th - 99th percentile	.938	.225	.489	1.387
Zscore production 18m	1st - 25th percentile	-.622	.223	-1.067	-.178
	26th - 49th percentile	-.153	.237	-.626	.320
	50th - 74th percentile	.163	.205	-.247	.573
	75th - 99th percentile	.894	.237	.421	1.368
Zscore gestures 18m	1st - 25th percentile	-.149	.243	-.634	.337
	26th - 49th percentile	.222	.259	-.295	.738
	50th - 74th percentile	-.212	.224	-.659	.236
	75th - 99th percentile	.229	.259	-.287	.746
Zscore Auditory Comprehension 18m	1st - 25th percentile	-.226	.242	-.710	.257
	26th - 49th percentile	.127	.257	-.388	.642
	50th - 74th percentile	-.158	.223	-.604	.288
	75th - 99th percentile	.340	.257	-.174	.855
Zscore Expressive Communication 18m	1st - 25th percentile	-.448	.221	-.889	-.007
	26th - 49th percentile	.087	.235	-.383	.557
	50th - 74th percentile	.005	.204	-.402	.412
	75th - 99th percentile	.616	.235	.146	1.086
Zscore comprehension 24m	1st - 25th percentile	-.557	.219	-.995	-.119
	26th - 49th percentile	-.130	.233	-.596	.336
	50th - 74th percentile	-.011	.202	-.415	.392
	75th - 99th percentile	.776	.233	.310	1.242
Zscore production 24m	1st - 25th percentile	-.442	.242	-.926	.041
	26th - 49th percentile	.240	.258	-.275	.755
	50th - 74th percentile	-.086	.223	-.531	.360
	75th - 99th percentile	.624	.258	.109	1.138
Zscore sentence Complexity 24m	1st - 25th percentile	-.526	.217	-.961	-.092
	26th - 49th percentile	.095	.231	-.368	.557
	50th - 74th percentile	.035	.200	-.365	.436

	75th - 99th percentile	.814	.231	.352	1.277
Zscore Auditory Comprehension 24m	1st - 25th percentile	-.106	.245	-.596	.384
	26th - 49th percentile	.215	.261	-.306	.737
	50th - 74th percentile	-.162	.226	-.614	.289
	75th - 99th percentile	.122	.261	-.400	.643
Zscore Expressive Communication 24m	1st - 25th percentile	-.393	.235	-.863	.077
	26th - 49th percentile	.285	.250	-.215	.785
	50th - 74th percentile	-.177	.217	-.610	.256
	75th - 99th percentile	.397	.250	-.103	.897
Zscore ASQ 36m	1st - 25th percentile	-.150	.211	-.572	.272
	26th - 49th percentile	-.183	.225	-.632	.266
	50th - 74th percentile	.027	.195	-.362	.416
	75th - 99th percentile	.293	.225	-.157	.742
Zscore production 36m	1st - 25th percentile	-.402	.221	-.845	.040
	26th - 49th percentile	.267	.236	-.204	.738
	50th - 74th percentile	-.170	.204	-.578	.238
	75th - 99th percentile	.525	.236	.054	.996
Zscore sentence Complexity 36m	1st - 25th percentile	-.687	.194	-1.074	-.299
	26th - 49th percentile	.215	.206	-.197	.628
	50th - 74th percentile	.078	.179	-.279	.436
	75th - 99th percentile	.469	.206	.056	.881

Multiple Comparisons

Tukey HSD

Dependent Variable	(I) production levels at 12m	(J) production levels at 12m	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval Lower Bound	Upper Bound
Zscore comprehension 12m	1st - 25th percentile	26th - 49th percentile	-.2707	.32202	.835	-1.1204	.5791
		50th - 74th percentile	-.3027	.29987	.744	-1.0941	.4886
		75th - 99th percentile	-1.4917*	.32202	.000	-2.3415	-.6419
	26th - 49th percentile	1st - 25th percentile	.2707	.32202	.835	-.5791	1.1204
		50th - 74th percentile	-.0321	.31049	1.000	-.8515	.7873
		75th - 99th percentile	-1.2210*	.33193	.003	-2.0970	-.3451
	50th - 74th percentile	1st - 25th percentile	.3027	.29987	.744	-.4886	1.0941
		26th - 49th percentile	.0321	.31049	1.000	-.7873	.8515
		75th - 99th percentile	-1.1889*	.31049	.002	-2.0083	-.3696
	75th - 99th percentile	1st - 25th percentile	1.4917*	.32202	.000	.6419	2.3415
		26th - 49th percentile	1.2210*	.33193	.003	.3451	2.0970
		50th - 74th percentile	1.1889*	.31049	.002	.3696	2.0083
Zscore gestures 12m		26th - 49th percentile	-.2437	.35104	.899	-1.1701	.6826

	1st - 25th percentile	50th - 74th percentile	-.0974	.32690	.991	-.9601	.7653
		75th - 99th percentile	-.6684	.35104	.237	-1.5948	.2580
	26th - 49th percentile	1st - 25th percentile	.2437	.35104	.899	-.6826	1.1701
		50th - 74th percentile	.1463	.33847	.973	-.7469	1.0395
		75th - 99th percentile	-.4247	.36184	.646	-1.3796	.5302
	50th - 74th percentile	1st - 25th percentile	.0974	.32690	.991	-.7653	.9601
		26th - 49th percentile	-.1463	.33847	.973	-1.0395	.7469
		75th - 99th percentile	-.5710	.33847	.339	-1.4642	.3222
	75th - 99th percentile	1st - 25th percentile	.6684	.35104	.237	-.2580	1.5948
		26th - 49th percentile	.4247	.36184	.646	-.5302	1.3796
		50th - 74th percentile	.5710	.33847	.339	-.3222	1.4642
Zscore comprehension 18m	1st - 25th percentile	26th - 49th percentile	-.1341	.30833	.972	-.9477	.6796
		50th - 74th percentile	-.3657	.28712	.583	-1.1234	.3920
		75th - 99th percentile	-1.3878*	.30833	.000	-2.2015	-.5741
	26th - 49th percentile	1st - 25th percentile	.1341	.30833	.972	-.6796	.9477
		50th - 74th percentile	-.2317	.29729	.864	-1.0162	.5529
		75th - 99th percentile	-1.2537*	.31782	.001	-2.0925	-.4150
	50th - 74th percentile	1st - 25th percentile	.3657	.28712	.583	-.3920	1.1234
		26th - 49th percentile	.2317	.29729	.864	-.5529	1.0162
		75th - 99th percentile	-1.0221*	.29729	.006	-1.8066	-.2375
	75th - 99th percentile	1st - 25th percentile	1.3878*	.30833	.000	.5741	2.2015
		26th - 49th percentile	1.2537*	.31782	.001	.4150	2.0925
		50th - 74th percentile	1.0221*	.29729	.006	.2375	1.8066
Zscore production 18m	1st - 25th percentile	26th - 49th percentile	-.4693	.32505	.477	-1.3271	.3885
		50th - 74th percentile	-.7848	.30270	.056	-1.5836	.0140
		75th - 99th percentile	-1.5165*	.32505	.000	-2.3743	-.6587
	26th - 49th percentile	1st - 25th percentile	.4693	.32505	.477	-.3885	1.3271
		50th - 74th percentile	-.3155	.31342	.746	-1.1426	.5116
		75th - 99th percentile	-1.0472*	.33506	.014	-1.9314	-.1630
	50th - 74th percentile	1st - 25th percentile	.7848	.30270	.056	-.0140	1.5836
		26th - 49th percentile	.3155	.31342	.746	-.5116	1.1426
		75th - 99th percentile	-.7317	.31342	.101	-1.5588	.0954
	75th - 99th percentile	1st - 25th percentile	1.5165*	.32505	.000	.6587	2.3743
		26th - 49th percentile	1.0472*	.33506	.014	.1630	1.9314
		50th - 74th percentile	.7317	.31342	.101	-.0954	1.5588
Zscore gestures 18m	1st - 25th percentile	26th - 49th percentile	-.3703	.35481	.725	-1.3066	.5661
		50th - 74th percentile	.0631	.33041	.998	-.8088	.9351
		75th - 99th percentile	-.3780	.35481	.712	-1.3143	.5583
	26th - 49th percentile	1st - 25th percentile	.3703	.35481	.725	-.5661	1.3066
		50th - 74th percentile	.4334	.34211	.587	-.4694	1.3362
		75th - 99th percentile	-.0078	.36573	1.000	-.9729	.9574
		1st - 25th percentile	-.0631	.33041	.998	-.9351	.8088

	50th - 74th percentile	26th - 49th percentile	-.4334	.34211	.587	-1.3362	.4694
		75th - 99th percentile	-.4411	.34211	.573	-1.3440	.4617
	75th - 99th percentile	1st - 25th percentile	.3780	.35481	.712	-.5583	1.3143
		26th - 49th percentile	.0078	.36573	1.000	-.9574	.9729
		50th - 74th percentile	.4411	.34211	.573	-.4617	1.3440
Zscore Auditory Comprehension 18m	1st - 25th percentile	26th - 49th percentile	-.3533	.35327	.750	-1.2856	.5789
		50th - 74th percentile	-.0683	.32897	.997	-.9364	.7998
		75th - 99th percentile	-.5665	.35327	.384	-1.4988	.3657
	26th - 49th percentile	1st - 25th percentile	.3533	.35327	.750	-.5789	1.2856
		50th - 74th percentile	.2850	.34062	.837	-.6138	1.1839
		75th - 99th percentile	-.2132	.36414	.936	-1.1741	.7477
	50th - 74th percentile	1st - 25th percentile	.0683	.32897	.997	-.7998	.9364
		26th - 49th percentile	-.2850	.34062	.837	-1.1839	.6138
		75th - 99th percentile	-.4982	.34062	.466	-1.3971	.4006
	75th - 99th percentile	1st - 25th percentile	.5665	.35327	.384	-.3657	1.4988
		26th - 49th percentile	.2132	.36414	.936	-.7477	1.1741
		50th - 74th percentile	.4982	.34062	.466	-.4006	1.3971
Zscore Expressive Communication 18m	1st - 25th percentile	26th - 49th percentile	-.5347	.32255	.355	-1.3859	.3165
		50th - 74th percentile	-.4527	.30037	.439	-1.2453	.3400
		75th - 99th percentile	-1.0639*	.32255	.008	-1.9151	-.2127
	26th - 49th percentile	1st - 25th percentile	.5347	.32255	.355	-.3165	1.3859
		50th - 74th percentile	.0820	.31101	.994	-.7387	.9027
		75th - 99th percentile	-.5292	.33248	.391	-1.4066	.3482
	50th - 74th percentile	1st - 25th percentile	.4527	.30037	.439	-.3400	1.2453
		26th - 49th percentile	-.0820	.31101	.994	-.9027	.7387
		75th - 99th percentile	-.6112	.31101	.212	-1.4319	.2096
	75th - 99th percentile	1st - 25th percentile	1.0639*	.32255	.008	.2127	1.9151
		26th - 49th percentile	.5292	.33248	.391	-.3482	1.4066
		50th - 74th percentile	.6112	.31101	.212	-.2096	1.4319
Zscore comprehension 24m	1st - 25th percentile	26th - 49th percentile	-.4272	.32007	.545	-1.2719	.4174
		50th - 74th percentile	-.5458	.29806	.269	-1.3323	.2408
		75th - 99th percentile	-1.3333*	.32007	.001	-2.1780	-.4887
	26th - 49th percentile	1st - 25th percentile	.4272	.32007	.545	-.4174	1.2719
		50th - 74th percentile	-.1186	.30861	.981	-.9330	.6958
		75th - 99th percentile	-.9061*	.32992	.038	-1.7768	-.0355
	50th - 74th percentile	1st - 25th percentile	.5458	.29806	.269	-.2408	1.3323
		26th - 49th percentile	.1186	.30861	.981	-.6958	.9330
		75th - 99th percentile	-.7876	.30861	.062	-1.6020	.0268
	75th - 99th percentile	1st - 25th percentile	1.3333*	.32007	.001	.4887	2.1780
		26th - 49th percentile	.9061*	.32992	.038	.0355	1.7768
		50th - 74th percentile	.7876	.30861	.062	-.0268	1.6020
		26th - 49th percentile	-.6825	.35332	.225	-1.6149	.2498

Zscore production 24m	1st - 25th percentile	50th - 74th percentile	-.3567	.32902	.701	-1.2249	.5116
		75th - 99th percentile	-1.0663*	.35332	.019	-1.9987	-.1339
	26th - 49th percentile	1st - 25th percentile	.6825	.35332	.225	-.2498	1.6149
		50th - 74th percentile	.3259	.34067	.774	-.5731	1.2249
		75th - 99th percentile	-.3838	.36419	.719	-1.3448	.5773
	50th - 74th percentile	1st - 25th percentile	.3567	.32902	.701	-.5116	1.2249
		26th - 49th percentile	-.3259	.34067	.774	-1.2249	.5731
		75th - 99th percentile	-.7096	.34067	.170	-1.6087	.1894
	75th - 99th percentile	1st - 25th percentile	1.0663*	.35332	.019	.1339	1.9987
		26th - 49th percentile	.3838	.36419	.719	-.5773	1.3448
		50th - 74th percentile	.7096	.34067	.170	-.1894	1.6087
Zscore sentence Complexity 24m	1st - 25th percentile	26th - 49th percentile	-.6214	.31754	.215	-1.4593	.2166
		50th - 74th percentile	-.5618	.29570	.239	-1.3422	.2185
		75th - 99th percentile	-1.3409*	.31754	.000	-2.1789	-.5029
	26th - 49th percentile	1st - 25th percentile	.6214	.31754	.215	-.2166	1.4593
		50th - 74th percentile	.0596	.30617	.997	-.7484	.8675
		75th - 99th percentile	-.7195	.32731	.135	-1.5833	.1442
	50th - 74th percentile	1st - 25th percentile	.5618	.29570	.239	-.2185	1.3422
		26th - 49th percentile	-.0596	.30617	.997	-.8675	.7484
		75th - 99th percentile	-.7791	.30617	.063	-1.5871	.0289
	75th - 99th percentile	1st - 25th percentile	1.3409*	.31754	.000	.5029	2.1789
		26th - 49th percentile	.7195	.32731	.135	-.1442	1.5833
		50th - 74th percentile	.7791	.30617	.063	-.0289	1.5871
Zscore Auditory Comprehension 24m	1st - 25th percentile	26th - 49th percentile	-.3213	.35806	.806	-1.2662	.6236
		50th - 74th percentile	.0563	.33344	.998	-.8236	.9362
		75th - 99th percentile	-.2276	.35806	.920	-1.1725	.7173
	26th - 49th percentile	1st - 25th percentile	.3213	.35806	.806	-.6236	1.2662
		50th - 74th percentile	.3776	.34525	.695	-.5335	1.2887
		75th - 99th percentile	.0937	.36908	.994	-.8803	1.0677
	50th - 74th percentile	1st - 25th percentile	-.0563	.33344	.998	-.9362	.8236
		26th - 49th percentile	-.3776	.34525	.695	-1.2887	.5335
		75th - 99th percentile	-.2839	.34525	.844	-1.1950	.6271
	75th - 99th percentile	1st - 25th percentile	.2276	.35806	.920	-.7173	1.1725
		26th - 49th percentile	-.0937	.36908	.994	-1.0677	.8803
		50th - 74th percentile	.2839	.34525	.844	-.6271	1.1950
Zscore Expressive Communication 24m	1st - 25th percentile	26th - 49th percentile	-.6782	.34329	.208	-1.5841	.2278
		50th - 74th percentile	-.2158	.31969	.906	-1.0595	.6278
		75th - 99th percentile	-.7898	.34329	.109	-1.6958	.1161
	26th - 49th percentile	1st - 25th percentile	.6782	.34329	.208	-.2278	1.5841
		50th - 74th percentile	.4623	.33101	.506	-.4112	1.3358
		75th - 99th percentile	-.1117	.35386	.989	-1.0455	.8222
		1st - 25th percentile	.2158	.31969	.906	-.6278	1.0595

	50th - 74th percentile	26th - 49th percentile	-.4623	.33101	.506	-1.3358	.4112
		75th - 99th percentile	-.5740	.33101	.315	-1.4475	.2995
	75th - 99th percentile	1st - 25th percentile	.7898	.34329	.109	-.1161	1.6958
		26th - 49th percentile	.1117	.35386	.989	-.8222	1.0455
		50th - 74th percentile	.5740	.33101	.315	-.2995	1.4475
Zscore ASQ 36m	1st - 25th percentile	26th - 49th percentile	.0333	.30840	1.000	-.7805	.8472
		50th - 74th percentile	-.1767	.28719	.927	-.9346	.5812
		75th - 99th percentile	-.4424	.30840	.483	-1.2562	.3715
	26th - 49th percentile	1st - 25th percentile	-.0333	.30840	1.000	-.8472	.7805
		50th - 74th percentile	-.2101	.29736	.894	-.9948	.5747
		75th - 99th percentile	-.4757	.31789	.446	-1.3146	.3632
	50th - 74th percentile	1st - 25th percentile	.1767	.28719	.927	-.5812	.9346
		26th - 49th percentile	.2101	.29736	.894	-.5747	.9948
		75th - 99th percentile	-.2657	.29736	.808	-1.0504	.5191
	75th - 99th percentile	1st - 25th percentile	.4424	.30840	.483	-.3715	1.2562
		26th - 49th percentile	.4757	.31789	.446	-.3632	1.3146
		50th - 74th percentile	.2657	.29736	.808	-.5191	1.0504
Zscore production 36m	1st - 25th percentile	26th - 49th percentile	-.6694	.32330	.174	-1.5225	.1838
		50th - 74th percentile	-.2328	.30107	.866	-1.0273	.5617
		75th - 99th percentile	-.9271*	.32330	.028	-1.7803	-.0739
	26th - 49th percentile	1st - 25th percentile	.6694	.32330	.174	-.1838	1.5225
		50th - 74th percentile	.4366	.31173	.504	-.3860	1.2592
		75th - 99th percentile	-.2577	.33325	.866	-1.1372	.6217
	50th - 74th percentile	1st - 25th percentile	.2328	.30107	.866	-.5617	1.0273
		26th - 49th percentile	-.4366	.31173	.504	-1.2592	.3860
		75th - 99th percentile	-.6943	.31173	.127	-1.5169	.1283
	75th - 99th percentile	1st - 25th percentile	.9271*	.32330	.028	.0739	1.7803
		26th - 49th percentile	.2577	.33325	.866	-.6217	1.1372
		50th - 74th percentile	.6943	.31173	.127	-.1283	1.5169
Zscore sentence Complexity 36m	1st - 25th percentile	26th - 49th percentile	-.9021*	.28314	.012	-1.6493	-.1549
		50th - 74th percentile	-.7652*	.26367	.026	-1.4610	-.0694
		75th - 99th percentile	-1.1555*	.28314	.001	-1.9027	-.4083
	26th - 49th percentile	1st - 25th percentile	.9021*	.28314	.012	.1549	1.6493
		50th - 74th percentile	.1370	.27301	.958	-.5835	.8574
		75th - 99th percentile	-.2533	.29186	.821	-1.0235	.5168
	50th - 74th percentile	1st - 25th percentile	.7652*	.26367	.026	.0694	1.4610
		26th - 49th percentile	-.1370	.27301	.958	-.8574	.5835
		75th - 99th percentile	-.3903	.27301	.486	-1.1108	.3301
	75th - 99th percentile	1st - 25th percentile	1.1555*	.28314	.001	.4083	1.9027
		26th - 49th percentile	.2533	.29186	.821	-.5168	1.0235
		50th - 74th percentile	.3903	.27301	.486	-.3301	1.1108

Based on observed means.

The error term is Mean Square(Error) = .639.

*. The mean difference is significant at the .05 level.

Production levels at 18m

Between-Subjects Factors

		Value Label	N
production levels at 18 months	1.00	1st - 25th percentile	15
	2.00	26th - 49th percentile	21
	3.00	50th - 74th percentile	16
	4.00	75th - 99th percentile	15

Descriptive Statistics

	production levels at 18m	Mean	Std. Deviation	N
Zscore comprehension 18m	1st - 25th percentile	-.9029	.59023	15
	26th - 49th percentile	-.2131	.76292	21
	50th - 74th percentile	.1641	.83144	16
	75th - 99th percentile	1.0262	.83257	15
	Total	.0000	1.00000	67
Zscore gestures 18m	1st - 25th percentile	-.7751	1.03009	15
	26th - 49th percentile	.0953	.75593	21
	50th - 74th percentile	.1140	.87543	16
	75th - 99th percentile	.5202	1.02647	15
	Total	.0000	1.00000	67
Zscore Auditory Comprehension 18m	1st - 25th percentile	-.6655	.81592	15
	26th - 49th percentile	-.2312	1.05422	21
	50th - 74th percentile	.2217	.52700	16
	75th - 99th percentile	.7527	.97230	15
	Total	.0000	1.00000	67
Zscore Expressive Communication 18m	1st - 25th percentile	-.8767	.72614	15
	26th - 49th percentile	-.2612	.54661	21
	50th - 74th percentile	.2796	.76787	16
	75th - 99th percentile	1.1456	.61792	15
	Total	.0451	.96382	67
Zscore comprehension 24m	1st - 25th percentile	-.8746	.65725	15
	26th - 49th percentile	-.3465	.87193	21
	50th - 74th percentile	.3133	.66701	16
	75th - 99th percentile	1.0256	.69339	15

	Total	.0000	1.00000	67
Zscore production 24m	1st - 25th percentile	-.7811	.77883	15
	26th - 49th percentile	-.4001	.79952	21
	50th - 74th percentile	.4074	.60238	16
	75th - 99th percentile	1.1546	.88479	15
	Total	.0555	1.04911	67
Zscore Sentence Complexity 24m	1st - 25th percentile	-.8494	.32634	15
	26th - 49th percentile	-.1385	.97960	21
	50th - 74th percentile	.3275	.38300	16
	75th - 99th percentile	1.0539	.96212	15
	Total	.0806	.99246	67
Zscore Auditory Comprehension 24m	1st - 25th percentile	-.3175	1.24362	15
	26th - 49th percentile	-.2766	1.00318	21
	50th - 74th percentile	.2888	.60922	16
	75th - 99th percentile	.3967	.91962	15
	Total	.0000	1.00000	67
Zscore Expressive Communication 24m	1st - 25th percentile	-.6152	.91759	15
	26th - 49th percentile	-.4497	.83905	21
	50th - 74th percentile	.4678	.67323	16
	75th - 99th percentile	.7457	.90902	15
	Total	.0000	1.00000	67
Zscore ASQ 36m	1st - 25th percentile	-.6346	.93814	15
	26th - 49th percentile	-.0641	.89292	21
	50th - 74th percentile	.3276	.63515	16
	75th - 99th percentile	.3507	.65046	15
	Total	-.0054	.86977	67
Zscore production 36m	1st - 25th percentile	-.5223	.96238	15
	26th - 49th percentile	-.3100	.87975	21
	50th - 74th percentile	.3041	.71236	16
	75th - 99th percentile	.7410	.80708	15
	Total	.0244	.96107	67
Zscore sentence Complexity 36m	1st - 25th percentile	-.6106	.98085	15
	26th - 49th percentile	-.0373	.90598	21
	50th - 74th percentile	.1572	.69465	16
	75th - 99th percentile	.5056	.62143	15
	Total	.0023	.89086	67

Box's Test of Equality of Covariance Matrices[a]

Box's M	425.319
F	1.150
df1	234
df2	7504.573
Sig.	.061

Tests the null hypothesis that the observed covariance matrices of the dependent variables are equal across groups.

a. Design: Intercept + production levels at 18m

Multivariate Tests[a]

Effect		Value	F	Hypothesis df	Error df	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power[d]
Intercept	Pillai's Trace	.047	.213[b]	12.000	52.000	.997	.047	2.551	.118
	Wilks' Lambda	.953	.213[b]	12.000	52.000	.997	.047	2.551	.118
	Hotelling's Trace	.049	.213[b]	12.000	52.000	.997	.047	2.551	.118
	Roy's Largest Root	.049	.213[b]	12.000	52.000	.997	.047	2.551	.118
production levels at 18m	Pillai's Trace	1.024	2.333	36.000	162.000	.000	.341	83.973	1.000
	Wilks' Lambda	.195	3.166	36.000	154.367	.000	.420	111.755	1.000
	Hotelling's Trace	3.059	4.305	36.000	152.000	.000	.505	154.981	1.000
	Roy's Largest Root	2.694	12.121[c]	12.000	54.000	.000	.729	145.455	1.000

a. Design: Intercept + production levels at 18m

b. Exact statistic

c. The statistic is an upper bound on F that yields a lower bound on the significance level.

d. Computed using alpha = .05

Levene's Test of Equality of Error Variances[a]

	F	df1	df2	Sig.
Zscore comprehension 18m	.433	3	63	.730
Zscore gestures 18m	.710	3	63	.550
Zscore Auditory Comprehension 18m	2.622	3	63	.058
Zscore Expressive Communication 18m	1.585	3	63	.202
Zscore comprehension 24m	1.018	3	63	.391
Zscore production 24m	.173	3	63	.914
Zscore sentence Complexity 24m	5.253	3	63	.003
Zscore Auditory Comprehension 24m	1.442	3	63	.239
Zscore Expressive communication 24m	.454	3	63	.715
Zscore ASQ 36m	.885	3	63	.454
Zscore production 36m	.540	3	63	.657
Zscore sentence Complexity 36m	2.365	3	63	.079

Tests the null hypothesis that the error variance of the dependent variable is equal across groups.

a. Design: Intercept + production levels at 18m

Tests of Between-Subjects Effects

Source	Dependent Variable	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power[m]
Corrected Model	Zscore comprehension 18m	29.408[a]	3	9.803	16.877	.000	.446	50.632	1.000
	Zscore gestures 18m	13.470[b]	3	4.490	5.385	.002	.204	16.154	.920
	Zscore Auditory Comprehension 18m	17.051[c]	3	5.684	7.315	.000	.258	21.946	.979
	Zscore Expressive Communication 18m	33.763[d]	3	11.254	25.738	.000	.551	77.213	1.000
	Zscore comprehension 24m	31.343[e]	3	10.448	18.991	.000	.475	56.974	1.000
	Zscore production 24m	34.962[f]	3	11.654	19.485	.000	.481	58.456	1.000
	Zscore sentence Complexity 24m	29.165[g]	3	9.722	17.088	.000	.449	51.263	1.000
	Zscore Auditory Comprehension 24m	6.813[h]	3	2.271	2.417	.075	.103	7.252	.577
	Zscore Expressive Communication 24m	21.765[i]	3	7.255	10.333	.000	.330	30.998	.998
	Zscore ASQ 36m	9.687[j]	3	3.229	5.055	.003	.194	15.165	.902
	Zscore production 36m	15.784[k]	3	5.261	7.337	.000	.259	22.010	.979
	Zscore sentence Complexity 36m	9.850[l]	3	3.283	4.863	.004	.188	14.590	.889
Intercept	Zscore comprehension 18m	.023	1	.023	.039	.844	.001	.039	.054
	Zscore gestures 18m	.009	1	.009	.010	.920	.000	.010	.051
	Zscore Auditory Comprehension 18m	.025	1	.025	.032	.859	.001	.032	.054

	Zscore Expressive Communication 18m	.339	1	.339	.775	.382	.012	.775	.140
	Zscore comprehension 24m	.057	1	.057	.103	.749	.002	.103	.062
	Zscore production 24m	.596	1	.596	.996	.322	.016	.996	.166
	Zscore sentence Complexity 24m	.636	1	.636	1.117	.295	.017	1.117	.180
	Zscore Auditory Comprehension 24m	.034	1	.034	.037	.849	.001	.037	.054
	Zscore Expressive Communication 24m	.091	1	.091	.129	.720	.002	.129	.064
	Zscore ASQ 36m	.002	1	.002	.003	.959	.000	.003	.050
	Zscore production 36m	.186	1	.186	.259	.612	.004	.259	.079
	Zscore sentence Complexity 36m	.001	1	.001	.001	.971	.000	.001	.050
production levels at 18m	Zscore comprehension 18m	29.408	3	9.803	16.877	.000	.446	50.632	1.000
	Zscore gestures 18m	13.470	3	4.490	5.385	.002	.204	16.154	.920
	Zscore Auditory Comprehension 18m	17.051	3	5.684	7.315	.000	.258	21.946	.979
	Zscore Expressive Communication 18m	33.763	3	11.254	25.738	.000	.551	77.213	1.000
	Zscore comprehension 24m	31.343	3	10.448	18.991	.000	.475	56.974	1.000
	Zscore production 24m	34.962	3	11.654	19.485	.000	.481	58.456	1.000
	Zscore sentence Complexity 24m	29.165	3	9.722	17.088	.000	.449	51.263	1.000
	Zscore Auditory Comprehension 24m	6.813	3	2.271	2.417	.075	.103	7.252	.577
	Zscore Expressive Communication 24m	21.765	3	7.255	10.333	.000	.330	30.998	.998
	Zscore ASQ 36m	9.687	3	3.229	5.055	.003	.194	15.165	.902
	Zscore production 36m	15.784	3	5.261	7.337	.000	.259	22.010	.979
	Zscore sentence Complexity 36m	9.850	3	3.283	4.863	.004	.188	14.590	.889
Error	Zscore comprehension 18m	36.592	63	.581					
	Zscore gestures 18m	52.530	63	.834					
	Zscore Auditory Comprehension 18m	48.949	63	.777					
	Zscore Expressive Communication 18m	27.548	63	.437					
	Zscore comprehension 24m	34.657	63	.550					
	Zscore production 24m	37.679	63	.598					
	Zscore sentence Complexity 24m	35.843	63	.569					

	Zscore Auditory Comprehension 24m	59.187	63	.939					
	Zscore Expressive Communication 24m	44.235	63	.702					
	Zscore ASQ 36m	40.242	63	.639					
	Zscore production 36m	45.177	63	.717					
	Zscore sentence Complexity 36m	42.530	63	.675					
Total	Zscore comprehension 18m	66.000	67						
	Zscore gestures 18m	66.000	67						
	Zscore Auditory Comprehension 18m	66.000	67						
	Zscore Expressive Communication 18m	61.446	67						
	Zscore comprehension 24m	66.000	67						
	Zscore production 24m	72.848	67						
	Zscore sentence Complexity 24m	65.443	67						
	Zscore Auditory Comprehension 24m	66.000	67						
	Zscore Expressive Communication 24m	66.000	67						
	Zscore ASQ 36m	49.931	67						
	Zscore production 36m	61.001	67						
	Zscore sentence Complexity 36m	52.380	67						
Corrected Total	Zscore comprehension 18m	66.000	66						
	Zscore gestures 18m	66.000	66						
	Zscore Auditory Comprehension 18m	66.000	66						
	Zscore Expressive Communication 18m	61.310	66						
	Zscore comprehension 24m	66.000	66						
	Zscore production 24m	72.641	66						
	Zscore sentence Complexity 24m	65.008	66						
	Zscore Auditory Comprehension 24m	66.000	66						
	Zscore Expressive Communication 24m	66.000	66						
	Zscore ASQ 36m	49.929	66						
	Zscore production 36m	60.961	66						

	Zscore sentence Complexity 36m	52.379	66							

a. R Squared = .446 (Adjusted R Squared = .419)
b. R Squared = .204 (Adjusted R Squared = .166)
c. R Squared = .258 (Adjusted R Squared = .223)
d. R Squared = .551 (Adjusted R Squared = .529)
e. R Squared = .475 (Adjusted R Squared = .450)
f. R Squared = .481 (Adjusted R Squared = .457)
g. R Squared = .449 (Adjusted R Squared = .422)
h. R Squared = .103 (Adjusted R Squared = .061)
i. R Squared = .330 (Adjusted R Squared = .298)
j. R Squared = .194 (Adjusted R Squared = .156)
k. R Squared = .259 (Adjusted R Squared = .224)
l. R Squared = .188 (Adjusted R Squared = .149)
m. Computed using alpha = .05

production levels at 18m

Dependent Variable	production levels at 18m	Mean	Std. Error	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
Zscore comprehension 18m	1st - 25th percentile	-.903	.197	-1.296	-.510
	26th - 49th percentile	-.213	.166	-.545	.119
	50th - 74th percentile	.164	.191	-.217	.545
	75th - 99th percentile	1.026	.197	.633	1.419
Zscore gestures 18m	1st - 25th percentile	-.775	.236	-1.246	-.304
	26th - 49th percentile	.095	.199	-.303	.493
	50th - 74th percentile	.114	.228	-.342	.570
	75th - 99th percentile	.520	.236	.049	.991
Zscore Auditory Comprehension 18m	1st - 25th percentile	-.666	.228	-1.120	-.211
	26th - 49th percentile	-.231	.192	-.616	.153
	50th - 74th percentile	.222	.220	-.219	.662
	75th - 99th percentile	.753	.228	.298	1.207
Zscore Expressive Communication 18m	1st - 25th percentile	-.877	.171	-1.218	-.536
	26th - 49th percentile	-.261	.144	-.550	.027
	50th - 74th percentile	.280	.165	-.051	.610
	75th - 99th percentile	1.146	.171	.804	1.487
Zscore comprehension 24m	1st - 25th percentile	-.875	.192	-1.257	-.492
	26th - 49th percentile	-.346	.162	-.670	-.023
	50th - 74th percentile	.313	.185	-.057	.684
	75th - 99th percentile	1.026	.192	.643	1.408

Zscore production 24m	1st - 25th percentile	-.781	.200	-1.180	-.382
	26th - 49th percentile	-.400	.169	-.737	-.063
	50th - 74th percentile	.407	.193	.021	.794
	75th - 99th percentile	1.155	.200	.756	1.554
Zscore sentence Complexity 24m	1st - 25th percentile	-.849	.195	-1.239	-.460
	26th - 49th percentile	-.139	.165	-.467	.190
	50th - 74th percentile	.327	.189	-.049	.704
	75th - 99th percentile	1.054	.195	.665	1.443
Zscore Auditory Comprehension 24m	1st - 25th percentile	-.318	.250	-.818	.183
	26th - 49th percentile	-.277	.212	-.699	.146
	50th - 74th percentile	.289	.242	-.195	.773
	75th - 99th percentile	.397	.250	-.103	.897
Zscore Expressive Communication 24m	1st - 25th percentile	-.615	.216	-1.048	-.183
	26th - 49th percentile	-.450	.183	-.815	-.084
	50th - 74th percentile	.468	.209	.049	.886
	75th - 99th percentile	.746	.216	.313	1.178
Zscore ASQ 36m	1st - 25th percentile	-.635	.206	-1.047	-.222
	26th - 49th percentile	-.064	.174	-.413	.284
	50th - 74th percentile	.328	.200	-.072	.727
	75th - 99th percentile	.351	.206	-.062	.763
Zscore production 36m	1st - 25th percentile	-.522	.219	-.959	-.085
	26th - 49th percentile	-.310	.185	-.679	.059
	50th - 74th percentile	.304	.212	-.119	.727
	75th - 99th percentile	.741	.219	.304	1.178
Zscore sentence Complexity 36m	1st - 25th percentile	-.611	.212	-1.034	-.187
	26th - 49th percentile	-.037	.179	-.396	.321
	50th - 74th percentile	.157	.205	-.253	.568
	75th - 99th percentile	.506	.212	.082	.929

Multiple Comparisons

Tukey HSD

Dependent Variable	(I) production levels at 18m	(J) production levels at 18m	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
Zscore comprehension 18m	1st - 25th percentile	26th - 49th percentile	-.6899*	.25764	.045	-1.3698	-.0100
		50th - 74th percentile	-1.0670*	.27390	.001	-1.7898	-.3442
		75th - 99th percentile	-1.9291*	.27829	.000	-2.6635	-1.1947

	26th - 49th percentile	1st - 25th percentile	.6899*	.25764	.045	.0100	1.3698
		50th - 74th percentile	-.3771	.25290	.449	-1.0445	.2903
		75th - 99th percentile	-1.2392*	.25764	.000	-1.9191	-.5593
	50th - 74th percentile	1st - 25th percentile	1.0670*	.27390	.001	.3442	1.7898
		26th - 49th percentile	.3771	.25290	.449	-.2903	1.0445
		75th - 99th percentile	-.8621*	.27390	.013	-1.5849	-.1393
	75th - 99th percentile	1st - 25th percentile	1.9291*	.27829	.000	1.1947	2.6635
		26th - 49th percentile	1.2392*	.25764	.000	.5593	1.9191
		50th - 74th percentile	.8621*	.27390	.013	.1393	1.5849
Zscore gestures 18m	1st - 25th percentile	26th - 49th percentile	-.8704*	.30870	.032	-1.6850	-.0557
		50th - 74th percentile	-.8891*	.32818	.042	-1.7551	-.0230
		75th - 99th percentile	-1.2953*	.33343	.001	-2.1752	-.4154
	26th - 49th percentile	1st - 25th percentile	.8704*	.30870	.032	.0557	1.6850
		50th - 74th percentile	-.0187	.30302	1.000	-.8183	.7809
		75th - 99th percentile	-.4249	.30870	.519	-1.2396	.3897
	50th - 74th percentile	1st - 25th percentile	.8891*	.32818	.042	.0230	1.7551
		26th - 49th percentile	.0187	.30302	1.000	-.7809	.8183
		75th - 99th percentile	-.4062	.32818	.605	-1.2723	.4598
	75th - 99th percentile	1st - 25th percentile	1.2953*	.33343	.001	.4154	2.1752
		26th - 49th percentile	.4249	.30870	.519	-.3897	1.2396
		50th - 74th percentile	.4062	.32818	.605	-.4598	1.2723

Zscore Auditory Comprehension 18m	1st - 25th percentile	26th - 49th percentile	-.4343	.29799	.469	-1.2207	.3520
		50th - 74th percentile	-.8873*	.31679	.033	-1.7233	-.0513
		75th - 99th percentile	-1.4182*	.32186	.000	-2.2676	-.5688
	26th - 49th percentile	1st - 25th percentile	.4343	.29799	.469	-.3520	1.2207
		50th - 74th percentile	-.4529	.29250	.415	-1.2248	.3190
		75th - 99th percentile	-.9839*	.29799	.008	-1.7703	-.1975
	50th - 74th percentile	1st - 25th percentile	.8873*	.31679	.033	.0513	1.7233
		26th - 49th percentile	.4529	.29250	.415	-.3190	1.2248
		75th - 99th percentile	-.5310	.31679	.345	-1.3670	.3050
	75th - 99th percentile	1st - 25th percentile	1.4182*	.32186	.000	.5688	2.2676
		26th - 49th percentile	.9839*	.29799	.008	.1975	1.7703
		50th - 74th percentile	.5310	.31679	.345	-.3050	1.3670
Zscore Expressive Communication 18m	1st - 25th percentile	26th - 49th percentile	-.6155*	.22355	.038	-1.2055	-.0256
		50th - 74th percentile	-1.1563*	.23766	.000	-1.7835	-.5291
		75th - 99th percentile	-2.0223*	.24146	.000	-2.6595	-1.3851
	26th - 49th percentile	1st - 25th percentile	.6155*	.22355	.038	.0256	1.2055
		50th - 74th percentile	-.5408	.21943	.076	-1.1199	.0383
		75th - 99th percentile	-1.4068*	.22355	.000	-1.9967	-.8169
	50th - 74th percentile	1st - 25th percentile	1.1563*	.23766	.000	.5291	1.7835
		26th - 49th percentile	.5408	.21943	.076	-.0383	1.1199
		75th - 99th percentile	-.8660*	.23766	.003	-1.4932	-.2389
	75th - 99th percentile	1st - 25th percentile	2.0223*	.24146	.000	1.3851	2.6595

		26th - 49th percentile	1.4068*	.22355	.000	.8169	1.9967
		50th - 74th percentile	.8660*	.23766	.003	.2389	1.4932
Zscore comprehension 24m	1st - 25th percentile	26th - 49th percentile	-.5281	.25074	.162	-1.1898	.1336
		50th - 74th percentile	-1.1879*	.26657	.000	-1.8914	-.4845
		75th - 99th percentile	-1.9002*	.27083	.000	-2.6149	-1.1855
	26th - 49th percentile	1st - 25th percentile	.5281	.25074	.162	-.1336	1.1898
		50th - 74th percentile	-.6598*	.24613	.045	-1.3093	-.0103
		75th - 99th percentile	-1.3721*	.25074	.000	-2.0337	-.7104
	50th - 74th percentile	1st - 25th percentile	1.1879*	.26657	.000	.4845	1.8914
		26th - 49th percentile	.6598*	.24613	.045	.0103	1.3093
		75th - 99th percentile	-.7123*	.26657	.046	-1.4157	-.0088
	75th - 99th percentile	1st - 25th percentile	1.9002*	.27083	.000	1.1855	2.6149
		26th - 49th percentile	1.3721*	.25074	.000	.7104	2.0337
		50th - 74th percentile	.7123*	.26657	.046	.0088	1.4157
Zscore production 24m	1st - 25th percentile	26th - 49th percentile	-.3810	.26144	.469	-1.0710	.3089
		50th - 74th percentile	-1.1886*	.27794	.000	-1.9221	-.4551
		75th - 99th percentile	-1.9358*	.28239	.000	-2.6810	-1.1906
	26th - 49th percentile	1st - 25th percentile	.3810	.26144	.469	-.3089	1.0710
		50th - 74th percentile	-.8075*	.25663	.013	-1.4848	-.1303
		75th - 99th percentile	-1.5547*	.26144	.000	-2.2447	-.8648
	50th - 74th percentile	1st - 25th percentile	1.1886*	.27794	.000	.4551	1.9221
		26th - 49th percentile	.8075*	.25663	.013	.1303	1.4848

		75th - 99th percentile	-.7472*	.27794	.044	-1.4807	-.0137
	75th - 99th percentile	1st - 25th percentile	1.9358*	.28239	.000	1.1906	2.6810
		26th - 49th percentile	1.5547*	.26144	.000	.8648	2.2447
		50th - 74th percentile	.7472*	.27794	.044	.0137	1.4807
Zscore sentence Complexity 24m	1st - 25th percentile	26th - 49th percentile	-.7108*	.25499	.034	-1.3837	-.0379
		50th - 74th percentile	-1.1768*	.27109	.000	-1.8922	-.4615
		75th - 99th percentile	-1.9032*	.27542	.000	-2.6301	-1.1764
	26th - 49th percentile	1st - 25th percentile	.7108*	.25499	.034	.0379	1.3837
		50th - 74th percentile	-.4660	.25030	.255	-1.1265	.1945
		75th - 99th percentile	-1.1924*	.25499	.000	-1.8653	-.5195
	50th - 74th percentile	1st - 25th percentile	1.1768*	.27109	.000	.4615	1.8922
		26th - 49th percentile	.4660	.25030	.255	-.1945	1.1265
		75th - 99th percentile	-.7264*	.27109	.045	-1.4418	-.0110
	75th - 99th percentile	1st - 25th percentile	1.9032*	.27542	.000	1.1764	2.6301
		26th - 49th percentile	1.1924*	.25499	.000	.5195	1.8653
		50th - 74th percentile	.7264*	.27109	.045	.0110	1.4418
Zscore Auditory Comprehension 24m	1st - 25th percentile	26th - 49th percentile	-.0410	.32767	.999	-.9057	.8237
		50th - 74th percentile	-.6063	.34835	.312	-1.5256	.3130
		75th - 99th percentile	-.7142	.35393	.192	-1.6482	.2198
	26th - 49th percentile	1st - 25th percentile	.0410	.32767	.999	-.8237	.9057
		50th - 74th percentile	-.5653	.32164	.303	-1.4141	.2835
		75th - 99th percentile	-.6733	.32767	.179	-1.5380	.1915

	50th - 74th percentile	1st - 25th percentile	.6063	.34835	.312	-.3130	1.5256
		26th - 49th percentile	.5653	.32164	.303	-.2835	1.4141
		75th - 99th percentile	-.1079	.34835	.990	-1.0272	.8113
	75th - 99th percentile	1st - 25th percentile	.7142	.35393	.192	-.2198	1.6482
		26th - 49th percentile	.6733	.32767	.179	-.1915	1.5380
		50th - 74th percentile	.1079	.34835	.990	-.8113	1.0272
Zscore Expressive Communication 24m	1st - 25th percentile	26th - 49th percentile	-.1655	.28328	.936	-.9130	.5821
		50th - 74th percentile	-1.0830*	.30115	.003	-1.8777	-.2883
		75th - 99th percentile	-1.3608*	.30597	.000	-2.1683	-.5534
	26th - 49th percentile	1st - 25th percentile	.1655	.28328	.936	-.5821	.9130
		50th - 74th percentile	-.9175*	.27806	.008	-1.6513	-.1837
		75th - 99th percentile	-1.1953*	.28328	.000	-1.9429	-.4478
	50th - 74th percentile	1st - 25th percentile	1.0830*	.30115	.003	.2883	1.8777
		26th - 49th percentile	.9175*	.27806	.008	.1837	1.6513
		75th - 99th percentile	-.2778	.30115	.793	-1.0726	.5169
	75th - 99th percentile	1st - 25th percentile	1.3608*	.30597	.000	.5534	2.1683
		26th - 49th percentile	1.1953*	.28328	.000	.4478	1.9429
		50th - 74th percentile	.2778	.30115	.793	-.5169	1.0726
Zscore ASQ 36m	1st - 25th percentile	26th - 49th percentile	-.5705	.27019	.161	-1.2835	.1425
		50th - 74th percentile	-.9622*	.28724	.007	-1.7202	-.2042
		75th - 99th percentile	-.9853*	.29184	.007	-1.7554	-.2151
	26th - 49th percentile	1st - 25th percentile	.5705	.27019	.161	-.1425	1.2835

		50th - 74th percentile	-.3917	.26522	.457	-1.0916	.3082
		75th - 99th percentile	-.4148	.27019	.423	-1.1278	.2982
	50th - 74th percentile	1st - 25th percentile	.9622*	.28724	.007	.2042	1.7202
		26th - 49th percentile	.3917	.26522	.457	-.3082	1.0916
		75th - 99th percentile	-.0231	.28724	1.000	-.7811	.7349
	75th - 99th percentile	1st - 25th percentile	.9853*	.29184	.007	.2151	1.7554
		26th - 49th percentile	.4148	.27019	.423	-.2982	1.1278
		50th - 74th percentile	.0231	.28724	1.000	-.7349	.7811
Zscore production 36m	1st - 25th percentile	26th - 49th percentile	-.2123	.28628	.880	-.9678	.5432
		50th - 74th percentile	-.8263*	.30434	.041	-1.6295	-.0232
		75th - 99th percentile	-1.2632*	.30921	.001	-2.0792	-.4472
	26th - 49th percentile	1st - 25th percentile	.2123	.28628	.880	-.5432	.9678
		50th - 74th percentile	-.6141	.28101	.139	-1.3556	.1275
		75th - 99th percentile	-1.0509*	.28628	.003	-1.8064	-.2955
	50th - 74th percentile	1st - 25th percentile	.8263*	.30434	.041	.0232	1.6295
		26th - 49th percentile	.6141	.28101	.139	-.1275	1.3556
		75th - 99th percentile	-.4369	.30434	.482	-1.2400	.3663
	75th - 99th percentile	1st - 25th percentile	1.2632*	.30921	.001	.4472	2.0792
		26th - 49th percentile	1.0509*	.28628	.003	.2955	1.8064
		50th - 74th percentile	.4369	.30434	.482	-.3663	1.2400
Zscore sentence Complexity 36m	1st - 25th percentile	26th - 49th percentile	-.5733	.27776	.176	-1.3063	.1597
		50th - 74th percentile	-.7677	.29529	.055	-1.5470	.0115

		75th - 99th percentile	-1.1161*	.30002	.002	-1.9078	-.3244
	26th - 49th percentile	1st - 25th percentile	.5733	.27776	.176	-.1597	1.3063
		50th - 74th percentile	-.1944	.27265	.892	-.9139	.5251
		75th - 99th percentile	-.5428	.27776	.216	-1.2758	.1902
	50th - 74th percentile	1st - 25th percentile	.7677	.29529	.055	-.0115	1.5470
		26th - 49th percentile	.1944	.27265	.892	-.5251	.9139
		75th - 99th percentile	-.3484	.29529	.642	-1.1276	.4309
	75th - 99th percentile	1st - 25th percentile	1.1161*	.30002	.002	.3244	1.9078
		26th - 49th percentile	.5428	.27776	.216	-.1902	1.2758
		50th - 74th percentile	.3484	.29529	.642	-.4309	1.1276

Based on observed means.

The error term is Mean Square(Error) = .675.

*. The mean difference is significant at the .05 level.

Comprehension levels at 12m

Between-Subjects Factors

		Value Label	N
Comprehension levels at 12m	1.00	1st - 25th percentile	17
	2.00	26th - 49th percentile	23
	3.00	50th - 74th percentile	17
	4.00	75th - 99th percentile	10

Descriptive Statistics

	Comprehension levels at 12m	Mean	Std. Deviation	N
Zscore production 12m	1st - 25th percentile	-.2155	.91456	17
	26th - 49th percentile	-.2159	.58619	23
	50th - 74th percentile	-.0336	.99081	17
	75th - 99th percentile	1.3804	1.10786	10

	Total	.0687	1.01738	67
Zscore gestures 12m	1st - 25th percentile	-.3042	.86007	17
	26th - 49th percentile	-.1163	.86528	23
	50th - 74th percentile	.1077	.85042	17
	75th - 99th percentile	.6015	1.50635	10
	Total	.0000	1.00000	67
Zscore comprehension 18m	1st - 25th percentile	-.5292	.51688	17
	26th - 49th percentile	-.4305	.80957	23
	50th - 74th percentile	.4714	.92668	17
	75th - 99th percentile	1.0884	.99548	10
	Total	.0000	1.00000	67
Zscore production 18m	1st - 25th percentile	-.0948	.80917	17
	26th - 49th percentile	-.2627	1.01313	23
	50th - 74th percentile	.0998	.85652	17
	75th - 99th percentile	.9746	1.34568	10
	Total	.0566	1.04654	67
Zscore gestures 18m	1st - 25th percentile	-.1555	.84046	17
	26th - 49th percentile	-.2104	1.00294	23
	50th - 74th percentile	.2243	.87572	17
	75th - 99th percentile	.3670	1.36081	10
	Total	.0000	1.00000	67
Zscore Auditory Comprehension 18m	1st - 25th percentile	-.0955	.64095	17
	26th - 49th percentile	-.4400	.95908	23
	50th - 74th percentile	.4730	1.12240	17
	75th - 99th percentile	.3703	1.01386	10
	Total	.0000	1.00000	67
Zscore Expressive Communication 18m	1st - 25th percentile	-.1782	.71072	17
	26th - 49th percentile	-.2601	1.02795	23
	50th - 74th percentile	.4449	1.02927	17
	75th - 99th percentile	.4468	.79655	10
	Total	.0451	.96382	67
Zscore comprehension 24m	1st - 25th percentile	-.2116	.77946	17
	26th - 49th percentile	-.6495	.73078	23
	50th - 74th percentile	.5633	.93999	17
	75th - 99th percentile	.8960	.85034	10
	Total	.0000	1.00000	67
Zscore production 24m	1st - 25th percentile	-.0143	.76432	17
	26th - 49th percentile	-.4924	.85242	23
	50th - 74th percentile	.4305	1.30771	17
	75th - 99th percentile	.7970	.75066	10
	Total	.0555	1.04911	67
	1st - 25th percentile	-.0582	1.02106	17

Zscore sentence Complexity 24m	26th - 49th percentile	-.1341	1.12451	23
	50th - 74th percentile	.2377	.95764	17
	75th - 99th percentile	.5429	.45830	10
	Total	.0806	.99246	67
Zscore Auditory Comprehension 24m	1st - 25th percentile	.0127	.65404	17
	26th - 49th percentile	-.3376	1.11125	23
	50th - 74th percentile	.3692	1.10990	17
	75th - 99th percentile	.1274	.89287	10
	Total	.0000	1.00000	67
Zscore Expressive Communication 24m	1st - 25th percentile	-.1406	.80064	17
	26th - 49th percentile	-.3363	1.03064	23
	50th - 74th percentile	.2412	1.16210	17
	75th - 99th percentile	.6026	.59717	10
	Total	.0000	1.00000	67
Zscore ASQ 36m	1st - 25th percentile	-.0630	.70004	17
	26th - 49th percentile	-.3604	.90645	23
	50th - 74th percentile	.3732	.83643	17
	75th - 99th percentile	.2654	.86233	10
	Total	-.0054	.86977	67
Zscore production 36m	1st - 25th percentile	-.0066	.66940	17
	26th - 49th percentile	-.3616	.96846	23
	50th - 74th percentile	.2487	1.08426	17
	75th - 99th percentile	.5839	.87670	10
	Total	.0244	.96107	67
Zscore sentence Complexity 36m	1st - 25th percentile	-.1698	.83792	17
	26th - 49th percentile	-.1077	.99913	23
	50th - 74th percentile	.0176	.90275	17
	75th - 99th percentile	.5223	.54308	10
	Total	.0023	.89086	67

Box's Test of Equality of Covariance Matrices[a]

Box's M	539.658
F	1.289
df1	240
df2	6665.406
Sig.	.002

Tests the null hypothesis that the observed covariance matrices of the dependent variables are equal across groups.

a. Design: Intercept + comprehension levels at 12m

Multivariate Tests[a]

Effect		Value	F	Hypothesis df	Error df	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power[d]
Intercept	Pillai's Trace	.110	.403[b]	15.000	49.000	.972	.110	6.042	.214
	Wilks' Lambda	.890	.403[b]	15.000	49.000	.972	.110	6.042	.214
	Hotelling's Trace	.123	.403[b]	15.000	49.000	.972	.110	6.042	.214
	Roy's Largest Root	.123	.403[b]	15.000	49.000	.972	.110	6.042	.214
comprehension levels at 12m	Pillai's Trace	1.081	1.914	45.000	153.000	.002	.360	86.129	1.000
	Wilks' Lambda	.229	2.089	45.000	146.347	.001	.388	92.837	1.000
	Hotelling's Trace	2.157	2.285	45.000	143.000	.000	.418	102.826	1.000
	Roy's Largest Root	1.517	5.159[c]	15.000	51.000	.000	.603	77.383	1.000

a. Design: Intercept + comprehension levels at 12m

b. Exact statistic

c. The statistic is an upper bound on F that yields a lower bound on the significance level.

d. Computed using alpha = .05

Levene's Test of Equality of Error Variances[a]

	F	df1	df2	Sig.
Zscore production 12m	3.395	3	63	.023
Zscore gestures 12m	2.819	3	63	.046
Zscore comprehension 18m	1.447	3	63	.238
Zscore production 18m	.841	3	63	.477
Zscore gestures 18m	.779	3	63	.510
Zscore Auditory Comprehension 18m	1.652	3	63	.187
Zscore Expressive Communication 18m	1.160	3	63	.332
Zscore comprehension 24m	1.080	3	63	.364
Zscore production 24m	1.348	3	63	.267
Zscore sentence Complexity 24m	1.808	3	63	.155
Zscore Auditory Comprehension 24m	1.445	3	63	.238
Zscore Expressive Communication 24m	.873	3	63	.460
Zscore ASQ 36m	.783	3	63	.508
Zscore production 36m	.882	3	63	.456
Zscore sentence Complexity 36m	2.058	3	63	.115

Tests the null hypothesis that the error variance of the dependent variable is equal across groups.

a. Design: Intercept + comprehension levels at 12m

Tests of Between-Subjects Effects

Source	Dependent Variable	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power[p]
Corrected Model	Zscore production 12m	20.619[a]	3	6.873	9.078	.000	.302	27.235	.994
	Zscore gestures 12m	5.700[b]	3	1.900	1.985	.125	.086	5.955	.488
	Zscore comprehension 18m	24.648[c]	3	8.216	12.517	.000	.373	37.551	1.000
	Zscore production 18m	11.193[d]	3	3.731	3.848	.014	.155	11.543	.798
	Zscore gestures 18m	3.632[e]	3	1.211	1.223	.309	.055	3.669	.312
	Zscore Auditory Comprehension 18m	9.783[f]	3	3.261	3.654	.017	.148	10.963	.775
	Zscore Expressive Communication 18m	7.320[g]	3	2.440	2.847	.045	.119	8.542	.656
	Zscore comprehension 24m	23.885[h]	3	7.962	11.910	.000	.362	35.730	.999
	Zscore production 24m	14.876[i]	3	4.959	5.408	.002	.205	16.224	.921

	Zscore sentence Complexity 24m	3.944[j]	3	1.315	1.356	.264	.061	4.069	.344
	Zscore Auditory Comprehension 24m	5.104[k]	3	1.701	1.760	.164	.077	5.280	.438
	Zscore Expressive Communication 24m	7.558[l]	3	2.519	2.716	.052	.115	8.147	.633
	Zscore ASQ 36m	6.125[m]	3	2.042	2.936	.040	.123	8.809	.671
	Zscore production 36m	7.430[n]	3	2.477	2.915	.041	.122	8.744	.667
	Zscore sentence Complexity 36m	3.490[o]	3	1.163	1.499	.223	.067	4.497	.378
Intercept	Zscore production 12m	3.209	1	3.209	4.239	.044	.063	4.239	.527
	Zscore gestures 12m	.319	1	.319	.334	.566	.005	.334	.088
	Zscore comprehension 18m	1.379	1	1.379	2.101	.152	.032	2.101	.298
	Zscore production 18m	1.968	1	1.968	2.030	.159	.031	2.030	.289
	Zscore gestures 18m	.195	1	.195	.197	.659	.003	.197	.072
	Zscore Auditory Comprehension 18m	.363	1	.363	.407	.526	.006	.407	.096
	Zscore Expressive Communication 18m	.787	1	.787	.919	.341	.014	.919	.157
	Zscore comprehension 24m	1.370	1	1.370	2.050	.157	.032	2.050	.292
	Zscore production 24m	1.990	1	1.990	2.170	.146	.033	2.170	.306
	Zscore sentence Complexity 24m	1.326	1	1.326	1.368	.247	.021	1.368	.210
	Zscore Auditory Comprehension 24m	.113	1	.113	.117	.734	.002	.117	.063
	Zscore Expressive Communication 24m	.515	1	.515	.556	.459	.009	.556	.114
	Zscore ASQ 36m	.177	1	.177	.255	.615	.004	.255	.079
	Zscore production 36m	.826	1	.826	.972	.328	.015	.972	.163
	Zscore sentence Complexity 36m	.264	1	.264	.340	.562	.005	.340	.089
comprehension levels at 12m	Zscore production 12m	20.619	3	6.873	9.078	.000	.302	27.235	.994
	Zscore gestures 12m	5.700	3	1.900	1.985	.125	.086	5.955	.488
	Zscore comprehension 18m	24.648	3	8.216	12.517	.000	.373	37.551	1.000
	Zscore production 18m	11.193	3	3.731	3.848	.014	.155	11.543	.798
	Zscore gestures 18m	3.632	3	1.211	1.223	.309	.055	3.669	.312
	Zscore Auditory Comprehension 18m	9.783	3	3.261	3.654	.017	.148	10.963	.775
	Zscore Expressive Communication 18m	7.320	3	2.440	2.847	.045	.119	8.542	.656
	Zscore comprehension 24m	23.885	3	7.962	11.910	.000	.362	35.730	.999
	Zscore production 24m	14.876	3	4.959	5.408	.002	.205	16.224	.921

	Zscore sentence Complexity 24m	3.944	3	1.315	1.356	.264	.061	4.069	.344
	Zscore Auditory Comprehension 24m	5.104	3	1.701	1.760	.164	.077	5.280	.438
	Zscore Expressive Communication 24m	7.558	3	2.519	2.716	.052	.115	8.147	.633
	Zscore ASQ 36m	6.125	3	2.042	2.936	.040	.123	8.809	.671
	Zscore production 36m	7.430	3	2.477	2.915	.041	.122	8.744	.667
	Zscore sentence Complexity 36m	3.490	3	1.163	1.499	.223	.067	4.497	.378
Error	Zscore production 12m	47.696	63	.757					
	Zscore gestures 12m	60.300	63	.957					
	Zscore comprehension 18m	41.352	63	.656					
	Zscore production 18m	61.093	63	.970					
	Zscore gestures 18m	62.368	63	.990					
	Zscore Auditory Comprehension 18m	56.217	63	.892					
	Zscore Expressive Communication 18m	53.990	63	.857					
	Zscore comprehension 24m	42.115	63	.668					
	Zscore production 24m	57.766	63	.917					
	Zscore sentence Complexity 24m	61.064	63	.969					
	Zscore Auditory Comprehension 24m	60.896	63	.967					
	Zscore Expressive Communication 24m	58.442	63	.928					
	Zscore ASQ 36m	43.804	63	.695					
	Zscore production 36m	53.531	63	.850					
	Zscore sentence Complexity 36m	48.889	63	.776					
Total	Zscore production 12m	68.631	67						
	Zscore gestures 12m	66.000	67						
	Zscore comprehension 18m	66.000	67						
	Zscore production 18m	72.501	67						
	Zscore gestures 18m	66.000	67						
	Zscore Auditory Comprehension 18m	66.000	67						
	Zscore Expressive Communication 18m	61.446	67						
	Zscore comprehension 24m	66.000	67						
	Zscore production 24m	72.848	67						

	Zscore sentence Complexity 24m	65.443	67							
	Zscore Auditory Comprehension 24m	66.000	67							
	Zscore Expressive Communication 24m	66.000	67							
	Zscore ASQ 36m	49.931	67							
	Zscore production 36m	61.001	67							
	Zscore sentence Complexity 36m	52.380	67							
Corrected Total	Zscore production 12m	68.314	66							
	Zscore gestures 12m	66.000	66							
	Zscore comprehension 18m	66.000	66							
	Zscore production 18m	72.287	66							
	Zscore gestures 18m	66.000	66							
	Zscore Auditory Comprehension 18m	66.000	66							
	Zscore Expressive Communication 18m	61.310	66							
	Zscore comprehension 24m	66.000	66							
	Zscore production 24m	72.641	66							
	Zscore sentence Complexity 24m	65.008	66							
	Zscore Auditory Comprehension 24m	66.000	66							
	Zscore Expressive Communication 24m	66.000	66							
	Zscore ASQ 36m	49.929	66							
	Zscore production 36m	60.961	66							
	Zscore sentence Complexity 36m	52.379	66							

a. R Squared = .302 (Adjusted R Squared = .269)

b. R Squared = .086 (Adjusted R Squared = .043)

c. R Squared = .373 (Adjusted R Squared = .344)

d. R Squared = .155 (Adjusted R Squared = .115)

e. R Squared = .055 (Adjusted R Squared = .010)

f. R Squared = .148 (Adjusted R Squared = .108)

g. R Squared = .119 (Adjusted R Squared = .077)

h. R Squared = .362 (Adjusted R Squared = .332)

i. R Squared = .205 (Adjusted R Squared = .167)

j. R Squared = .061 (Adjusted R Squared = .016)

k. R Squared = .077 (Adjusted R Squared = .033)

l. R Squared = .115 (Adjusted R Squared = .072)

m. R Squared = .123 (Adjusted R Squared = .081)

n. R Squared = .122 (Adjusted R Squared = .080)

o. R Squared = .067 (Adjusted R Squared = .022)

p. Computed using alpha = .05

Comprehension levels at 12m

Dependent Variable	Comprehension levels at 12m	Mean	Std. Error	95% Confidence Interval	
				Lower Bound	Upper Bound
Zscore production 12m	1st - 25th percentile	-.216	.211	-.637	.206
	26th - 49th percentile	-.216	.181	-.578	.147
	50th - 74th percentile	-.034	.211	-.455	.388
	75th - 99th percentile	1.380	.275	.831	1.930
Zscore gestures 12m	1st - 25th percentile	-.304	.237	-.778	.170
	26th - 49th percentile	-.116	.204	-.524	.291
	50th - 74th percentile	.108	.237	-.366	.582
	75th - 99th percentile	.602	.309	-.017	1.220
Zscore comprehension 18m	1st - 25th percentile	-.529	.196	-.922	-.136
	26th - 49th percentile	-.431	.169	-.768	-.093
	50th - 74th percentile	.471	.196	.079	.864
	75th - 99th percentile	1.088	.256	.576	1.600
Zscore production 18m	1st - 25th percentile	-.095	.239	-.572	.383
	26th - 49th percentile	-.263	.205	-.673	.148
	50th - 74th percentile	.100	.239	-.377	.577
	75th - 99th percentile	.975	.311	.352	1.597
Zscore gestures 18m	1st - 25th percentile	-.156	.241	-.638	.327
	26th - 49th percentile	-.210	.207	-.625	.204
	50th - 74th percentile	.224	.241	-.258	.707
	75th - 99th percentile	.367	.315	-.262	.996
Zscore Auditory Comprehension 18m	1st - 25th percentile	-.095	.229	-.553	.362
	26th - 49th percentile	-.440	.197	-.834	-.046
	50th - 74th percentile	.473	.229	.015	.931
	75th - 99th percentile	.370	.299	-.227	.967
Zscore Expressive Communication 18m	1st - 25th percentile	-.178	.225	-.627	.270
	26th - 49th percentile	-.260	.193	-.646	.126
	50th - 74th percentile	.445	.225	-.004	.894
	75th - 99th percentile	.447	.293	-.138	1.032
Zscore comprehension 24m	1st - 25th percentile	-.212	.198	-.608	.185
	26th - 49th percentile	-.649	.170	-.990	-.309
	50th - 74th percentile	.563	.198	.167	.960
	75th - 99th percentile	.896	.259	.379	1.413
Zscore production 24m	1st - 25th percentile	-.014	.232	-.478	.450

	26th - 49th percentile	-.492	.200	-.891	-.093
	50th - 74th percentile	.430	.232	-.034	.895
	75th - 99th percentile	.797	.303	.192	1.402
Zscore sentence Complexity 24m	1st - 25th percentile	-.058	.239	-.535	.419
	26th - 49th percentile	-.134	.205	-.544	.276
	50th - 74th percentile	.238	.239	-.239	.715
	75th - 99th percentile	.543	.311	-.079	1.165
Zscore Auditory Comprehension 24m	1st - 25th percentile	.013	.238	-.464	.489
	26th - 49th percentile	-.338	.205	-.747	.072
	50th - 74th percentile	.369	.238	-.107	.846
	75th - 99th percentile	.127	.311	-.494	.749
Zscore Expressive Communication 24m	1st - 25th percentile	-.141	.234	-.607	.326
	26th - 49th percentile	-.336	.201	-.738	.065
	50th - 74th percentile	.241	.234	-.226	.708
	75th - 99th percentile	.603	.305	-.006	1.211
Zscore ASQ 36m	1st - 25th percentile	-.063	.202	-.467	.341
	26th - 49th percentile	-.360	.174	-.708	-.013
	50th - 74th percentile	.373	.202	-.031	.777
	75th - 99th percentile	.265	.264	-.261	.792
Zscore production 36m	1st - 25th percentile	-.007	.224	-.453	.440
	26th - 49th percentile	-.362	.192	-.746	.022
	50th - 74th percentile	.249	.224	-.198	.695
	75th - 99th percentile	.584	.291	.001	1.166
Zscore sentence Complexity 36m	1st - 25th percentile	-.170	.214	-.597	.257
	26th - 49th percentile	-.108	.184	-.475	.259
	50th - 74th percentile	.018	.214	-.409	.445
	75th - 99th percentile	.522	.279	-.034	1.079

Multiple Comparisons

Tukey HSD

Dependent Variable	(I) Comprehension levels at 12m	(J) Comprehension vels at 12m	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
Zscore production 12m	1st - 25th percentile	26th - 49th percentile	.0004	.27830	1.000	-.7340	.7348
		50th - 74th percentile	-.1819	.29844	.929	-.9695	.6056
		75th - 99th percentile	-1.5959*	.34676	.000	-2.5110	-.6808
	26th - 49th percentile	1st - 25th percentile	-.0004	.27830	1.000	-.7348	.7340
		50th - 74th percentile	-.1823	.27830	.913	-.9168	.5521
		75th - 99th percentile	-1.5963*	.32958	.000	-2.4660	-.7265
	50th - 74th percentile	1st - 25th percentile	.1819	.29844	.929	-.6056	.9695

		26th - 49th percentile	.1823	.27830	.913	-.5521	.9168
		75th - 99th percentile	-1.4139*	.34676	.001	-2.3290	-.4989
	75th - 99th percentile	1st - 25th percentile	1.5959*	.34676	.000	.6808	2.5110
		26th - 49th percentile	1.5963*	.32958	.000	.7265	2.4660
		50th - 74th percentile	1.4139*	.34676	.001	.4989	2.3290
Zscore gestures 12m	1st - 25th percentile	26th - 49th percentile	-.1879	.31292	.932	-1.0136	.6379
		50th - 74th percentile	-.4118	.33557	.612	-1.2974	.4737
		75th - 99th percentile	-.9057	.38989	.104	-1.9346	.1232
	26th - 49th percentile	1st - 25th percentile	.1879	.31292	.932	-.6379	1.0136
		50th - 74th percentile	-.2240	.31292	.890	-1.0498	.6018
		75th - 99th percentile	-.7178	.37058	.223	-1.6958	.2601
	50th - 74th percentile	1st - 25th percentile	.4118	.33557	.612	-.4737	1.2974
		26th - 49th percentile	.2240	.31292	.890	-.6018	1.0498
		75th - 99th percentile	-.4939	.38989	.587	-1.5228	.5350
	75th - 99th percentile	1st - 25th percentile	.9057	.38989	.104	-.1232	1.9346
		26th - 49th percentile	.7178	.37058	.223	-.2601	1.6958
		50th - 74th percentile	.4939	.38989	.587	-.5350	1.5228
Zscore comprehension 18m	1st - 25th percentile	26th - 49th percentile	-.0986	.25913	.981	-.7824	.5852
		50th - 74th percentile	-1.0006*	.27789	.003	-1.7339	-.2672
		75th - 99th percentile	-1.6176*	.32288	.000	-2.4696	-.7655
	26th - 49th percentile	1st - 25th percentile	.0986	.25913	.981	-.5852	.7824
		50th - 74th percentile	-.9019*	.25913	.005	-1.5858	-.2181
		75th - 99th percentile	-1.5190*	.30688	.000	-2.3288	-.7091
	50th - 74th percentile	1st - 25th percentile	1.0006*	.27789	.003	.2672	1.7339
		26th - 49th percentile	.9019*	.25913	.005	.2181	1.5858
		75th - 99th percentile	-.6170	.32288	.234	-1.4691	.2350
	75th - 99th percentile	1st - 25th percentile	1.6176*	.32288	.000	.7655	2.4696
		26th - 49th percentile	1.5190*	.30688	.000	.7091	2.3288
		50th - 74th percentile	.6170	.32288	.234	-.2350	1.4691
Zscore production 18m	1st - 25th percentile	26th - 49th percentile	.1679	.31497	.951	-.6633	.9991
		50th - 74th percentile	-.1946	.33777	.939	-1.0859	.6968
		75th - 99th percentile	-1.0694*	.39245	.040	-2.1050	-.0337
	26th - 49th percentile	1st - 25th percentile	-.1679	.31497	.951	-.9991	.6633
		50th - 74th percentile	-.3625	.31497	.660	-1.1937	.4687
		75th - 99th percentile	-1.2373*	.37301	.008	-2.2216	-.2529
	50th - 74th percentile	1st - 25th percentile	.1946	.33777	.939	-.6968	1.0859
		26th - 49th percentile	.3625	.31497	.660	-.4687	1.1937
		75th - 99th percentile	-.8748	.39245	.127	-1.9104	.1609
	75th - 99th percentile	1st - 25th percentile	1.0694*	.39245	.040	.0337	2.1050
		26th - 49th percentile	1.2373*	.37301	.008	.2529	2.2216
		50th - 74th percentile	.8748	.39245	.127	-.1609	1.9104
Zscore gestures 18m	1st - 25th percentile	26th - 49th percentile	.0549	.31824	.998	-.7849	.8947

		50th - 74th percentile	-.3798	.34127	.683	-1.2804	.5208
		75th - 99th percentile	-.5225	.39652	.555	-1.5689	.5239
	26th - 49th percentile	1st - 25th percentile	-.0549	.31824	.998	-.8947	.7849
		50th - 74th percentile	-.4347	.31824	.525	-1.2746	.4051
		75th - 99th percentile	-.5774	.37688	.425	-1.5720	.4171
	50th - 74th percentile	1st - 25th percentile	.3798	.34127	.683	-.5208	1.2804
		26th - 49th percentile	.4347	.31824	.525	-.4051	1.2746
		75th - 99th percentile	-.1427	.39652	.984	-1.1891	.9037
	75th - 99th percentile	1st - 25th percentile	.5225	.39652	.555	-.5239	1.5689
		26th - 49th percentile	.5774	.37688	.425	-.4171	1.5720
		50th - 74th percentile	.1427	.39652	.984	-.9037	1.1891
Zscore Auditory Comprehension 18m	1st - 25th percentile	26th - 49th percentile	.3446	.30214	.666	-.4527	1.1419
		50th - 74th percentile	-.5684	.32401	.305	-1.4235	.2866
		75th - 99th percentile	-.4658	.37646	.606	-1.4593	.5277
	26th - 49th percentile	1st - 25th percentile	-.3446	.30214	.666	-1.1419	.4527
		50th - 74th percentile	-.9130*	.30214	.019	-1.7103	-.1157
		75th - 99th percentile	-.8104	.35781	.117	-1.7546	.1339
	50th - 74th percentile	1st - 25th percentile	.5684	.32401	.305	-.2866	1.4235
		26th - 49th percentile	.9130*	.30214	.019	.1157	1.7103
		75th - 99th percentile	.1026	.37646	.993	-.8908	1.0961
	75th - 99th percentile	1st - 25th percentile	.4658	.37646	.606	-.5277	1.4593
		26th - 49th percentile	.8104	.35781	.117	-.1339	1.7546
		50th - 74th percentile	-.1026	.37646	.993	-1.0961	.8908
Zscore Expressive Communication 18m	1st - 25th percentile	26th - 49th percentile	.0819	.29609	.993	-.6995	.8633
		50th - 74th percentile	-.6231	.31752	.213	-1.4610	.2148
		75th - 99th percentile	-.6249	.36893	.335	-1.5985	.3487
	26th - 49th percentile	1st - 25th percentile	-.0819	.29609	.993	-.8633	.6995
		50th - 74th percentile	-.7050	.29609	.091	-1.4864	.0764
		75th - 99th percentile	-.7068	.35065	.193	-1.6322	.2185
	50th - 74th percentile	1st - 25th percentile	.6231	.31752	.213	-.2148	1.4610
		26th - 49th percentile	.7050	.29609	.091	-.0764	1.4864
		75th - 99th percentile	-.0018	.36893	1.000	-.9754	.9718
	75th - 99th percentile	1st - 25th percentile	.6249	.36893	.335	-.3487	1.5985
		26th - 49th percentile	.7068	.35065	.193	-.2185	1.6322
		50th - 74th percentile	.0018	.36893	1.000	-.9718	.9754
Zscore comprehension 24m	1st - 25th percentile	26th - 49th percentile	.4379	.26151	.346	-.2522	1.1280
		50th - 74th percentile	-.7749*	.28044	.037	-1.5149	-.0348
		75th - 99th percentile	-1.1076*	.32584	.006	-1.9675	-.2477
	26th - 49th percentile	1st - 25th percentile	-.4379	.26151	.346	-1.1280	.2522
		50th - 74th percentile	-1.2128*	.26151	.000	-1.9029	-.5226
		75th - 99th percentile	-1.5455*	.30970	.000	-2.3628	-.7282
	50th - 74th percentile	1st - 25th percentile	.7749*	.28044	.037	.0348	1.5149

		26th - 49th percentile	1.2128*	.26151	.000	.5226	1.9029
		75th - 99th percentile	-.3327	.32584	.738	-1.1926	.5272
	75th - 99th percentile	1st - 25th percentile	1.1076*	.32584	.006	.2477	1.9675
		26th - 49th percentile	1.5455*	.30970	.000	.7282	2.3628
		50th - 74th percentile	.3327	.32584	.738	-.5272	1.1926
Zscore production 24m	1st - 25th percentile	26th - 49th percentile	.4780	.30627	.408	-.3302	1.2863
		50th - 74th percentile	-.4448	.32844	.532	-1.3115	.4219
		75th - 99th percentile	-.8114	.38161	.156	-1.8184	.1957
	26th - 49th percentile	1st - 25th percentile	-.4780	.30627	.408	-1.2863	.3302
		50th - 74th percentile	-.9228*	.30627	.019	-1.7311	-.1146
		75th - 99th percentile	-1.2894*	.36271	.004	-2.2466	-.3322
	50th - 74th percentile	1st - 25th percentile	.4448	.32844	.532	-.4219	1.3115
		26th - 49th percentile	.9228*	.30627	.019	.1146	1.7311
		75th - 99th percentile	-.3666	.38161	.772	-1.3736	.6405
	75th - 99th percentile	1st - 25th percentile	.8114	.38161	.156	-.1957	1.8184
		26th - 49th percentile	1.2894*	.36271	.004	.3322	2.2466
		50th - 74th percentile	.3666	.38161	.772	-.6405	1.3736
Zscore sentence Complexity 24m	1st - 25th percentile	26th - 49th percentile	.0759	.31489	.995	-.7551	.9069
		50th - 74th percentile	-.2959	.33769	.817	-1.1870	.5952
		75th - 99th percentile	-.6010	.39236	.425	-1.6364	.4344
	26th - 49th percentile	1st - 25th percentile	-.0759	.31489	.995	-.9069	.7551
		50th - 74th percentile	-.3718	.31489	.641	-1.2028	.4592
		75th - 99th percentile	-.6770	.37292	.276	-1.6611	.3072
	50th - 74th percentile	1st - 25th percentile	.2959	.33769	.817	-.5952	1.1870
		26th - 49th percentile	.3718	.31489	.641	-.4592	1.2028
		75th - 99th percentile	-.3051	.39236	.864	-1.3405	.7303
	75th - 99th percentile	1st - 25th percentile	.6010	.39236	.425	-.4344	1.6364
		26th - 49th percentile	.6770	.37292	.276	-.3072	1.6611
		50th - 74th percentile	.3051	.39236	.864	-.7303	1.3405
Zscore Auditory Comprehension 24m	1st - 25th percentile	26th - 49th percentile	.3504	.31446	.682	-.4795	1.1802
		50th - 74th percentile	-.3564	.33722	.717	-1.2463	.5335
		75th - 99th percentile	-.1147	.39182	.991	-1.1487	.9193
	26th - 49th percentile	1st - 25th percentile	-.3504	.31446	.682	-1.1802	.4795
		50th - 74th percentile	-.7068	.31446	.122	-1.5366	.1231
		75th - 99th percentile	-.4650	.37241	.599	-1.4478	.5177
	50th - 74th percentile	1st - 25th percentile	.3564	.33722	.717	-.5335	1.2463
		26th - 49th percentile	.7068	.31446	.122	-.1231	1.5366
		75th - 99th percentile	.2418	.39182	.926	-.7922	1.2757
	75th - 99th percentile	1st - 25th percentile	.1147	.39182	.991	-.9193	1.1487
		26th - 49th percentile	.4650	.37241	.599	-.5177	1.4478
		50th - 74th percentile	-.2418	.39182	.926	-1.2757	.7922
	1st - 25th percentile	26th - 49th percentile	.1957	.30806	.920	-.6172	1.0087

Zscore Expressive Communication 24m		50th - 74th percentile	-.3818	.33036	.657	-1.2536	.4900
		75th - 99th percentile	-.7432	.38384	.224	-1.7562	.2697
	26th - 49th percentile	1st - 25th percentile	-.1957	.30806	.920	-1.0087	.6172
		50th - 74th percentile	-.5775	.30806	.249	-1.3904	.2355
		75th - 99th percentile	-.9389	.36483	.059	-1.9017	.0238
	50th - 74th percentile	1st - 25th percentile	.3818	.33036	.657	-.4900	1.2536
		26th - 49th percentile	.5775	.30806	.249	-.2355	1.3904
		75th - 99th percentile	-.3615	.38384	.783	-1.3744	.6515
	75th - 99th percentile	1st - 25th percentile	.7432	.38384	.224	-.2697	1.7562
		26th - 49th percentile	.9389	.36483	.059	-.0238	1.9017
		50th - 74th percentile	.3615	.38384	.783	-.6515	1.3744
Zscore ASQ 36m	1st - 25th percentile	26th - 49th percentile	.2973	.26670	.682	-.4065	1.0012
		50th - 74th percentile	-.4362	.28601	.429	-1.1910	.3185
		75th - 99th percentile	-.3285	.33231	.757	-1.2054	.5485
	26th - 49th percentile	1st - 25th percentile	-.2973	.26670	.682	-1.0012	.4065
		50th - 74th percentile	-.7336*	.26670	.038	-1.4374	-.0298
		75th - 99th percentile	-.6258	.31585	.206	-1.4593	.2077
	50th - 74th percentile	1st - 25th percentile	.4362	.28601	.429	-.3185	1.1910
		26th - 49th percentile	.7336*	.26670	.038	.0298	1.4374
		75th - 99th percentile	.1078	.33231	.988	-.7692	.9847
	75th - 99th percentile	1st - 25th percentile	.3285	.33231	.757	-.5485	1.2054
		26th - 49th percentile	.6258	.31585	.206	-.2077	1.4593
		50th - 74th percentile	-.1078	.33231	.988	-.9847	.7692
Zscore production 36m	1st - 25th percentile	26th - 49th percentile	.3550	.29483	.627	-.4231	1.1330
		50th - 74th percentile	-.2553	.31617	.851	-1.0897	.5791
		75th - 99th percentile	-.5906	.36736	.382	-1.5600	.3789
	26th - 49th percentile	1st - 25th percentile	-.3550	.29483	.627	-1.1330	.4231
		50th - 74th percentile	-.6103	.29483	.174	-1.3883	.1678
		75th - 99th percentile	-.9456*	.34916	.042	-1.8670	-.0241
	50th - 74th percentile	1st - 25th percentile	.2553	.31617	.851	-.5791	1.0897
		26th - 49th percentile	.6103	.29483	.174	-.1678	1.3883
		75th - 99th percentile	-.3353	.36736	.798	-1.3047	.6342
	75th - 99th percentile	1st - 25th percentile	.5906	.36736	.382	-.3789	1.5600
		26th - 49th percentile	.9456*	.34916	.042	.0241	1.8670
		50th - 74th percentile	.3353	.36736	.798	-.6342	1.3047
Zscore sentence Complexity 36m	1st - 25th percentile	26th - 49th percentile	-.0621	.28176	.996	-.8056	.6815
		50th - 74th percentile	-.1874	.30215	.925	-.9848	.6100
		75th - 99th percentile	-.6921	.35107	.210	-1.6186	.2343
	26th - 49th percentile	1st - 25th percentile	.0621	.28176	.996	-.6815	.8056
		50th - 74th percentile	-.1253	.28176	.970	-.8689	.6182
		75th - 99th percentile	-.6300	.33368	.244	-1.5106	.2505
	50th - 74th percentile	1st - 25th percentile	.1874	.30215	.925	-.6100	.9848

		26th - 49th percentile	.1253	.28176	.970	-.6182	.8689
		75th - 99th percentile	-.5047	.35107	.481	-1.4312	.4217
	75th - 99th percentile	1st - 25th percentile	.6921	.35107	.210	-.2343	1.6186
		26th - 49th percentile	.6300	.33368	.244	-.2505	1.5106
		50th - 74th percentile	.5047	.35107	.481	-.4217	1.4312

Based on observed means.

The error term is Mean Square(Error) = .776.

*. The mean difference is significant at the .05 level.

Comprehension levels at 18m

Between-Subjects Factors

		Value Label	N
comprehension levels at 18m	1.00	1st - 25th percentile	22
	2.00	26th - 49th percentile	16
	3.00	50th - 74th percentile	15
	4.00	75th - 99th percentile	14

Descriptive Statistics

	comprehension levels at 18m	Mean	Std. Deviation	N
Zscore production 18m	1st - 25th percentile	-.5332	.82514	22
	26th - 49th percentile	-.2967	.78472	16
	50th - 74th percentile	.2877	.72451	15
	75th - 99th percentile	1.1395	1.05728	14
	Total	.0566	1.04654	67
Zscore gestures 18m	1st - 25th percentile	-.8684	.87900	22
	26th - 49th percentile	.2558	.68608	16
	50th - 74th percentile	.3689	.70629	15
	75th - 99th percentile	.6770	.86094	14
	Total	.0000	1.00000	67

Zscore Auditory Comprehension 18m	1st - 25th percentile	-.7148	.66437	22
	26th - 49th percentile	-.1042	.94401	16
	50th - 74th percentile	.3402	.79199	15
	75th - 99th percentile	.8778	.90480	14
	Total	.0000	1.00000	67
Zscore Expressive Communication 18m	1st - 25th percentile	-.3499	.90103	22
	26th - 49th percentile	-.3635	.60313	16
	50th - 74th percentile	.2632	.89940	15
	75th - 99th percentile	.8991	.89664	14
	Total	.0451	.96382	67
Zscore comprehension 24m	1st - 25th percentile	-.7129	.73580	22
	26th - 49th percentile	-.4407	.66576	16
	50th - 74th percentile	.3842	.73663	15
	75th - 99th percentile	1.2123	.50429	14
	Total	.0000	1.00000	67
Zscore production 24m	1st - 25th percentile	-.4041	.76536	22
	26th - 49th percentile	-.4237	.68246	16
	50th - 74th percentile	.0815	.94800	15
	75th - 99th percentile	1.2977	.91307	14
	Total	.0555	1.04911	67
Zscore sentence Complexity 24m	1st - 25th percentile	-.2757	.81541	22
	26th - 49th percentile	-.3610	.70424	16
	50th - 74th percentile	.2310	.97931	15

	75th - 99th percentile	.9839	.98073	14
	Total	.0806	.99246	67
Zscore Auditory Comprehension 24m	1st - 25th percentile	-.6198	.96499	22
	26th - 49th percentile	.1296	.87750	16
	50th - 74th percentile	.3323	.81604	15
	75th - 99th percentile	.4699	.96490	14
	Total	.0000	1.00000	67
Zscore Expressive Communication 24m	1st - 25th percentile	-.2453	.77059	22
	26th - 49th percentile	-.3631	.89876	16
	50th - 74th percentile	-.0150	1.11851	15
	75th - 99th percentile	.8165	.92529	14
	Total	.0000	1.00000	67
Zscore ASQ 36m	1st - 25th percentile	-.2173	.84975	22
	26th - 49th percentile	-.2281	.91802	16
	50th - 74th percentile	.0905	.97987	15
	75th - 99th percentile	.4794	.51687	14
	Total	-.0054	.86977	67
Zscore production 36m	1st - 25th percentile	-.4288	.80033	22
	26th - 49th percentile	-.2057	.72023	16
	50th - 74th percentile	-.1067	.96878	15
	75th - 99th percentile	1.1401	.52183	14
	Total	.0244	.96107	67
Zscore sentence Complexity 36m	1st - 25th percentile	-.3513	.87418	22

	26th - 49th percentile	-.1022	.80140	16
	50th - 74th percentile	.0963	1.02096	15
	75th - 99th percentile	.5769	.58336	14
	Total	.0023	.89086	67

Box's Test of Equality of Covariance Matrices[a]

Box's M	405.415
F	1.085
df1	234
df2	7039.855
Sig.	.182

Tests the null hypothesis that the observed covariance matrices of the dependent variables are equal across groups.

a. Design: Intercept + comprehension levels at 18m

Multivariate Tests[a]

Effect		Value	F	Hypothesis df	Error df	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power[d]
Intercept	Pillai's Trace	.089	.423[b]	12.000	52.000	.948	.089	5.071	.208
	Wilks' Lambda	.911	.423[b]	12.000	52.000	.948	.089	5.071	.208
	Hotelling's Trace	.098	.423[b]	12.000	52.000	.948	.089	5.071	.208
	Roy's Largest Root	.098	.423[b]	12.000	52.000	.948	.089	5.071	.208
comprehension levels at 18m	Pillai's Trace	1.163	2.848	36.000	162.000	.000	.388	102.542	1.000
	Wilks' Lambda	.158	3.715	36.000	154.367	.000	.459	131.033	1.000
	Hotelling's Trace	3.430	4.828	36.000	152.000	.000	.533	173.807	1.000
	Roy's Largest Root	2.830	12.733[c]	12.000	54.000	.000	.739	152.801	1.000

a. Design: Intercept + comprehension levels at 18m

b. Exact statistic

c. The statistic is an upper bound on F that yields a lower bound on the significance level.

d. Computed using alpha = .05

Levene's Test of Equality of Error Variances[a]

	F	df1	df2	Sig.
Zscore production 18m	.569	3	63	.637
Zscore gestures 18m	.506	3	63	.679
Zscore Auditory Comprehension 18m	1.770	3	63	.162
Zscore Expressive Communication 18m	1.850	3	63	.147
Zscore comprehension 24m	.675	3	63	.571
Zscore production 24m	.714	3	63	.547
Zscore sentence Complexity 24m	.809	3	63	.494
Zscore Auditory Comprehension 24m	.123	3	63	.946
Zscore Expressive Communication 24m	.773	3	63	.514
Zscore ASQ 36m	1.459	3	63	.234
Zscore production 36m	.865	3	63	.464
Zscore sentence Complexity 36m	1.575	3	63	.204

Tests the null hypothesis that the error variance of the dependent variable is equal across groups.

a. Design: Intercept + comprehension levels at 18m

Tests of Between-Subjects Effects

Source	Dependent Variable	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power[m]
Corrected Model	Zscore production 18m	26.871[a]	3	8.957	12.425	.000	.372	37.276	1.000
	Zscore gestures 18m	26.094[b]	3	8.698	13.732	.000	.395	41.196	1.000
	Zscore Auditory Comprehension 18m	23.939[c]	3	7.980	11.952	.000	.363	35.857	.999
	Zscore Expressive Communication 18m	17.028[d]	3	5.676	8.075	.000	.278	24.226	.988
	Zscore comprehension 24m	37.079[e]	3	12.360	26.924	.000	.562	80.773	1.000
	Zscore production 24m	29.934[f]	3	9.978	14.719	.000	.412	44.157	1.000
	Zscore sentence Complexity 24m	17.676[g]	3	5.892	7.842	.000	.272	23.526	.985
	Zscore Auditory Comprehension 24m	13.468[h]	3	4.489	5.384	.002	.204	16.152	.920
	Zscore Expressive Communication 24m	12.768[i]	3	4.256	5.037	.003	.193	15.112	.901
	Zscore ASQ 36m	5.209[j]	3	1.736	2.446	.072	.104	7.338	.583
	Zscore production 36m	23.049[k]	3	7.683	12.767	.000	.378	38.302	1.000

	Zscore sentence Complexity 36m	7.681[l]	3	2.560	3.608	.018	.147	10.825	.769
Intercept	Zscore production 18m	1.450	1	1.450	2.011	.161	.031	2.011	.287
	Zscore gestures 18m	.763	1	.763	1.205	.277	.019	1.205	.191
	Zscore Auditory Comprehension 18m	.647	1	.647	.969	.329	.015	.969	.163
	Zscore Expressive Communication 18m	.819	1	.819	1.165	.285	.018	1.165	.186
	Zscore comprehension 24m	.797	1	.797	1.737	.192	.027	1.737	.254
	Zscore production 24m	1.235	1	1.235	1.822	.182	.028	1.822	.265
	Zscore sentence Complexity 24m	1.359	1	1.359	1.808	.184	.028	1.808	.263
	Zscore Auditory Comprehension 24m	.395	1	.395	.474	.494	.007	.474	.104
	Zscore Expressive Communication 24m	.152	1	.152	.179	.673	.003	.179	.070
	Zscore ASQ 36m	.063	1	.063	.089	.767	.001	.089	.060
	Zscore production 36m	.647	1	.647	1.075	.304	.017	1.075	.175
	Zscore sentence Complexity 36m	.196	1	.196	.276	.601	.004	.276	.081
comprehension levels at 18m	Zscore production 18m	26.871	3	8.957	12.425	.000	.372	37.276	1.000
	Zscore gestures 18m	26.094	3	8.698	13.732	.000	.395	41.196	1.000
	Zscore Auditory Comprehension 18m	23.939	3	7.980	11.952	.000	.363	35.857	.999
	Zscore Expressive Communication 18m	17.028	3	5.676	8.075	.000	.278	24.226	.988
	Zscore comprehension 24m	37.079	3	12.360	26.924	.000	.562	80.773	1.000
	Zscore production 24m	29.934	3	9.978	14.719	.000	.412	44.157	1.000
	Zscore sentence Complexity 24m	17.676	3	5.892	7.842	.000	.272	23.526	.985
	Zscore Auditory Comprehension 24m	13.468	3	4.489	5.384	.002	.204	16.152	.920
	Zscore Expressive Communication 24m	12.768	3	4.256	5.037	.003	.193	15.112	.901
	Zscore ASQ 36m	5.209	3	1.736	2.446	.072	.104	7.338	.583
	Zscore production 36m	23.049	3	7.683	12.767	.000	.378	38.302	1.000
	Zscore sentence Complexity 36m	7.681	3	2.560	3.608	.018	.147	10.825	.769
Error	Zscore production 18m	45.415	63	.721					

	Zscore gestures 18m	39.906	63	.633						
	Zscore Auditory Comprehension 18m	42.061	63	.668						
	Zscore Expressive Communication 18m	44.282	63	.703						
	Zscore comprehension 24m	28.921	63	.459						
	Zscore production 24m	42.708	63	.678						
	Zscore sentence Complexity 24m	47.333	63	.751						
	Zscore Auditory Comprehension 24m	52.532	63	.834						
	Zscore Expressive Communication 24m	53.232	63	.845						
	Zscore ASQ 36m	44.720	63	.710						
	Zscore production 36m	37.912	63	.602						
	Zscore sentence Complexity 36m	44.699	63	.710						
Total	Zscore production 18m	72.501	67							
	Zscore gestures 18m	66.000	67							
	Zscore Auditory Comprehension 18m	66.000	67							
	Zscore Expressive Communication 18m	61.446	67							
	Zscore comprehension 24m	66.000	67							
	Zscore production 24m	72.848	67							
	Zscore sentence Complexity 24m	65.443	67							
	Zscore Auditory Comprehension 24m	66.000	67							
	Zscore Expressive Communication 24m	66.000	67							
	Zscore ASQ 36m	49.931	67							
	Zscore production 36m	61.001	67							
	Zscore sentence Complexity 36m	52.380	67							
Corrected Total	Zscore production 18m	72.287	66							
	Zscore gestures 18m	66.000	66							
	Zscore Auditory Comprehension 18m	66.000	66							

	Zscore Expressive Communication 18m	61.310	66							
	Zscore comprehension 24m	66.000	66							
	Zscore production 24m	72.641	66							
	Zscore sentence Complexity 24m	65.008	66							
	Zscore Auditory Comprehension 24m	66.000	66							
	Zscore Expressive Communication 24m	66.000	66							
	Zscore ASQ 36m	49.929	66							
	Zscore production 36m	60.961	66							
	Zscore sentence Complexity 36m	52.379	66							

a. R Squared = .372 (Adjusted R Squared = .342)
b. R Squared = .395 (Adjusted R Squared = .367)
c. R Squared = .363 (Adjusted R Squared = .332)
d. R Squared = .278 (Adjusted R Squared = .243)
e. R Squared = .562 (Adjusted R Squared = .541)
f. R Squared = .412 (Adjusted R Squared = .384)
g. R Squared = .272 (Adjusted R Squared = .237)
h. R Squared = .204 (Adjusted R Squared = .166)
i. R Squared = .193 (Adjusted R Squared = .155)
j. R Squared = .104 (Adjusted R Squared = .062)
k. R Squared = .378 (Adjusted R Squared = .348)
l. R Squared = .147 (Adjusted R Squared = .106)
m. Computed using alpha = .05

Estimated Marginal Means

comprehension levels at 18m

Dependent Variable	comprehension levels at 18m	Mean	Std. Error	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
Zscore production 18m	1st - 25th percentile	-.533	.181	-.895	-.172
	26th - 49th percentile	-.297	.212	-.721	.127
	50th - 74th percentile	.288	.219	-.150	.726
	75th - 99th percentile	1.140	.227	.686	1.593
Zscore gestures 18m	1st - 25th percentile	-.868	.170	-1.207	-.529
	26th - 49th percentile	.256	.199	-.142	.653
	50th - 74th percentile	.369	.205	-.042	.780
	75th - 99th percentile	.677	.213	.252	1.102

Zscore Auditory Comprehension 18m	1st - 25th percentile	-.715	.174	-1.063	-.367
	26th - 49th percentile	-.104	.204	-.512	.304
	50th - 74th percentile	.340	.211	-.081	.762
	75th - 99th percentile	.878	.218	.441	1.314
Zscore Expressive Communication 18m	1st - 25th percentile	-.350	.179	-.707	.007
	26th - 49th percentile	-.364	.210	-.782	.055
	50th - 74th percentile	.263	.216	-.169	.696
	75th - 99th percentile	.899	.224	.451	1.347
Zscore comprehension 24m	1st - 25th percentile	-.713	.144	-1.002	-.424
	26th - 49th percentile	-.441	.169	-.779	-.102
	50th - 74th percentile	.384	.175	.035	.734
	75th - 99th percentile	1.212	.181	.850	1.574
Zscore production 24m	1st - 25th percentile	-.404	.176	-.755	-.053
	26th - 49th percentile	-.424	.206	-.835	-.012
	50th - 74th percentile	.082	.213	-.343	.506
	75th - 99th percentile	1.298	.220	.858	1.737
Zscore sentence Complexity 24m	1st - 25th percentile	-.276	.185	-.645	.094
	26th - 49th percentile	-.361	.217	-.794	.072
	50th - 74th percentile	.231	.224	-.216	.678
	75th - 99th percentile	.984	.232	.521	1.447
Zscore Auditory Comprehension 24m	1st - 25th percentile	-.620	.195	-1.009	-.231
	26th - 49th percentile	.130	.228	-.327	.586
	50th - 74th percentile	.332	.236	-.139	.803
	75th - 99th percentile	.470	.244	-.018	.958
Zscore Expressive Communication 24m	1st - 25th percentile	-.245	.196	-.637	.146
	26th - 49th percentile	-.363	.230	-.822	.096
	50th - 74th percentile	-.015	.237	-.489	.459
	75th - 99th percentile	.816	.246	.326	1.307
Zscore ASQ 36m	1st - 25th percentile	-.217	.180	-.576	.142
	26th - 49th percentile	-.228	.211	-.649	.193
	50th - 74th percentile	.091	.218	-.344	.525
	75th - 99th percentile	.479	.225	.029	.929
Zscore production 36m	1st - 25th percentile	-.429	.165	-.759	-.098
	26th - 49th percentile	-.206	.194	-.593	.182
	50th - 74th percentile	-.107	.200	-.507	.294
	75th - 99th percentile	1.140	.207	.726	1.554
Zscore sentence Complexity 36m	1st - 25th percentile	-.351	.180	-.710	.008
	26th - 49th percentile	-.102	.211	-.523	.319
	50th - 74th percentile	.096	.217	-.338	.531
	75th - 99th percentile	.577	.225	.127	1.027

Multiple Comparisons

Tukey HSD

Dependent Variable	(I) comprehension levels at 18m	(J) comprehension levels at 18m	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
Zscore production 18m	1st - 25th percentile	26th - 49th percentile	-.2365	.27897	.831	-.9727	.4997
		50th - 74th percentile	-.8209*	.28430	.027	-1.5712	-.0707
		75th - 99th percentile	-1.6728*	.29027	.000	-2.4388	-.9068
	26th - 49th percentile	1st - 25th percentile	.2365	.27897	.831	-.4997	.9727
		50th - 74th percentile	-.5844	.30515	.232	-1.3897	.2208
		75th - 99th percentile	-1.4363*	.31072	.000	-2.2562	-.6163
	50th - 74th percentile	1st - 25th percentile	.8209*	.28430	.027	.0707	1.5712
		26th - 49th percentile	.5844	.30515	.232	-.2208	1.3897
		75th - 99th percentile	-.8518*	.31552	.043	-1.6845	-.0192
	75th - 99th percentile	1st - 25th percentile	1.6728*	.29027	.000	.9068	2.4388
		26th - 49th percentile	1.4363*	.31072	.000	.6163	2.2562
		50th - 74th percentile	.8518*	.31552	.043	.0192	1.6845
Zscore gestures 18m	1st - 25th percentile	26th - 49th percentile	-1.1241*	.26150	.000	-1.8142	-.4340
		50th - 74th percentile	-1.2373*	.26650	.000	-1.9406	-.5340
		75th - 99th percentile	-1.5454*	.27210	.000	-2.2634	-.8273
	26th - 49th percentile	1st - 25th percentile	1.1241*	.26150	.000	.4340	1.8142
		50th - 74th percentile	-.1132	.28604	.979	-.8680	.6416
		75th - 99th percentile	-.4212	.29126	.476	-1.1899	.3474
	50th - 74th percentile	1st - 25th percentile	1.2373*	.26650	.000	.5340	1.9406
		26th - 49th percentile	.1132	.28604	.979	-.6416	.8680
		75th - 99th percentile	-.3080	.29576	.726	-1.0885	.4725
	75th - 99th percentile	1st - 25th percentile	1.5454*	.27210	.000	.8273	2.2634
		26th - 49th percentile	.4212	.29126	.476	-.3474	1.1899
		50th - 74th percentile	.3080	.29576	.726	-.4725	1.0885
Zscore Auditory Comprehension 18m	1st - 25th percentile	26th - 49th percentile	-.6107	.26847	.115	-1.3191	.0978
		50th - 74th percentile	-1.0550*	.27360	.002	-1.7770	-.3330
		75th - 99th percentile	-1.5927*	.27935	.000	-2.3298	-.8555
	26th - 49th percentile	1st - 25th percentile	.6107	.26847	.115	-.0978	1.3191
		50th - 74th percentile	-.4444	.29366	.436	-1.2193	.3306
		75th - 99th percentile	-.9820*	.29902	.009	-1.7711	-.1929

	50th - 74th percentile	1st - 25th percentile	1.0550*	.27360	.002	.3330	1.7770
		26th - 49th percentile	.4444	.29366	.436	-.3306	1.2193
		75th - 99th percentile	-.5376	.30364	.297	-1.3389	.2637
	75th - 99th percentile	1st - 25th percentile	1.5927*	.27935	.000	.8555	2.3298
		26th - 49th percentile	.9820*	.29902	.009	.1929	1.7711
		50th - 74th percentile	.5376	.30364	.297	-.2637	1.3389
Zscore Expressive Communication 18m	1st - 25th percentile	26th - 49th percentile	.0136	.27546	1.000	-.7133	.7406
		50th - 74th percentile	-.6132	.28073	.139	-1.3540	.1277
		75th - 99th percentile	-1.2490*	.28663	.000	-2.0054	-.4926
	26th - 49th percentile	1st - 25th percentile	-.0136	.27546	1.000	-.7406	.7133
		50th - 74th percentile	-.6268	.30131	.171	-1.4219	.1684
		75th - 99th percentile	-1.2626*	.30682	.001	-2.0723	-.4529
	50th - 74th percentile	1st - 25th percentile	.6132	.28073	.139	-.1277	1.3540
		26th - 49th percentile	.6268	.30131	.171	-.1684	1.4219
		75th - 99th percentile	-.6358	.31155	.184	-1.4580	.1864
	75th - 99th percentile	1st - 25th percentile	1.2490*	.28663	.000	.4926	2.0054
		26th - 49th percentile	1.2626*	.30682	.001	.4529	2.0723
		50th - 74th percentile	.6358	.31155	.184	-.1864	1.4580
Zscore comprehension 24m	1st - 25th percentile	26th - 49th percentile	-.2722	.22261	.615	-.8597	.3152
		50th - 74th percentile	-1.0972*	.22687	.000	-1.6959	-.4985
		75th - 99th percentile	-1.9252*	.23164	.000	-2.5365	-1.3140
	26th - 49th percentile	1st - 25th percentile	.2722	.22261	.615	-.3152	.8597
		50th - 74th percentile	-.8249*	.24351	.007	-1.4675	-.1823
		75th - 99th percentile	-1.6530*	.24795	.000	-2.3073	-.9987
	50th - 74th percentile	1st - 25th percentile	1.0972*	.22687	.000	.4985	1.6959
		26th - 49th percentile	.8249*	.24351	.007	.1823	1.4675
		75th - 99th percentile	-.8281*	.25178	.009	-1.4925	-.1636
	75th - 99th percentile	1st - 25th percentile	1.9252*	.23164	.000	1.3140	2.5365
		26th - 49th percentile	1.6530*	.24795	.000	.9987	2.3073
		50th - 74th percentile	.8281*	.25178	.009	.1636	1.4925
Zscore production 24m	1st - 25th percentile	26th - 49th percentile	.0196	.27052	1.000	-.6943	.7335
		50th - 74th percentile	-.4856	.27569	.301	-1.2132	.2419
		75th - 99th percentile	-1.7018*	.28149	.000	-2.4446	-.9589
	26th - 49th percentile	1st - 25th percentile	-.0196	.27052	1.000	-.7335	.6943
		50th - 74th percentile	-.5052	.29591	.328	-1.2861	.2757
		75th - 99th percentile	-1.7214*	.30131	.000	-2.5165	-.9262
	50th - 74th percentile	1st - 25th percentile	.4856	.27569	.301	-.2419	1.2132
		26th - 49th percentile	.5052	.29591	.328	-.2757	1.2861
		75th - 99th percentile	-1.2162*	.30596	.001	-2.0236	-.4087
	75th - 99th percentile	1st - 25th percentile	1.7018*	.28149	.000	.9589	2.4446
		26th - 49th percentile	1.7214*	.30131	.000	.9262	2.5165
		50th - 74th percentile	1.2162*	.30596	.001	.4087	2.0236

Zscore sentence Complexity 24m	1st - 25th percentile	26th - 49th percentile	.0852	.28479	.991	-.6663	.8368
		50th - 74th percentile	-.5068	.29024	.309	-1.2727	.2592
		75th - 99th percentile	-1.2596*	.29634	.000	-2.0416	-.4776
	26th - 49th percentile	1st - 25th percentile	-.0852	.28479	.991	-.8368	.6663
		50th - 74th percentile	-.5920	.31152	.238	-1.4141	.2301
		75th - 99th percentile	-1.3449*	.31721	.000	-2.1820	-.5078
	50th - 74th percentile	1st - 25th percentile	.5068	.29024	.309	-.2592	1.2727
		26th - 49th percentile	.5920	.31152	.238	-.2301	1.4141
		75th - 99th percentile	-.7529	.32211	.100	-1.6029	.0972
	75th - 99th percentile	1st - 25th percentile	1.2596*	.29634	.000	.4776	2.0416
		26th - 49th percentile	1.3449*	.31721	.000	.5078	2.1820
		50th - 74th percentile	.7529	.32211	.100	-.0972	1.6029
Zscore Auditory Comprehension 24m	1st - 25th percentile	26th - 49th percentile	-.7494	.30003	.070	-1.5412	.0423
		50th - 74th percentile	-.9521*	.30576	.014	-1.7590	-.1452
		75th - 99th percentile	-1.0897*	.31219	.005	-1.9136	-.2659
	26th - 49th percentile	1st - 25th percentile	.7494	.30003	.070	-.0423	1.5412
		50th - 74th percentile	-.2027	.32818	.926	-1.0688	.6633
		75th - 99th percentile	-.3403	.33418	.739	-1.2222	.5416
	50th - 74th percentile	1st - 25th percentile	.9521*	.30576	.014	.1452	1.7590
		26th - 49th percentile	.2027	.32818	.926	-.6633	1.0688
		75th - 99th percentile	-.1376	.33934	.977	-1.0331	.7579
	75th - 99th percentile	1st - 25th percentile	1.0897*	.31219	.005	.2659	1.9136
		26th - 49th percentile	.3403	.33418	.739	-.5416	1.2222
		50th - 74th percentile	.1376	.33934	.977	-.7579	1.0331
Zscore Expressive Communication 24m	1st - 25th percentile	26th - 49th percentile	.1178	.30202	.980	-.6793	.9148
		50th - 74th percentile	-.2303	.30779	.877	-1.0425	.5820
		75th - 99th percentile	-1.0617*	.31426	.007	-1.8911	-.2324
	26th - 49th percentile	1st - 25th percentile	-.1178	.30202	.980	-.9148	.6793
		50th - 74th percentile	-.3481	.33036	.719	-1.2199	.5237
		75th - 99th percentile	-1.1795*	.33640	.005	-2.0672	-.2918
	50th - 74th percentile	1st - 25th percentile	.2303	.30779	.877	-.5820	1.0425
		26th - 49th percentile	.3481	.33036	.719	-.5237	1.2199
		75th - 99th percentile	-.8315	.34159	.081	-1.7329	.0700
	75th - 99th percentile	1st - 25th percentile	1.0617*	.31426	.007	.2324	1.8911
		26th - 49th percentile	1.1795*	.33640	.005	.2918	2.0672
		50th - 74th percentile	.8315	.34159	.081	-.0700	1.7329
Zscore ASQ 36m	1st - 25th percentile	26th - 49th percentile	.0108	.27682	1.000	-.7197	.7414
		50th - 74th percentile	-.3078	.28211	.696	-1.0523	.4367
		75th - 99th percentile	-.6966	.28804	.084	-1.4568	.0635
	26th - 49th percentile	1st - 25th percentile	-.0108	.27682	1.000	-.7414	.7197
		50th - 74th percentile	-.3186	.30280	.720	-1.1177	.4805
		75th - 99th percentile	-.7075	.30833	.110	-1.5211	.1062

	50th - 74th percentile	1st - 25th percentile	.3078	.28211	.696	-.4367	1.0523
		26th - 49th percentile	.3186	.30280	.720	-.4805	1.1177
		75th - 99th percentile	-.3888	.31309	.603	-1.2151	.4374
	75th - 99th percentile	1st - 25th percentile	.6966	.28804	.084	-.0635	1.4568
		26th - 49th percentile	.7075	.30833	.110	-.1062	1.5211
		50th - 74th percentile	.3888	.31309	.603	-.4374	1.2151
Zscore production 36m	1st - 25th percentile	26th - 49th percentile	-.2231	.25488	.818	-.8957	.4495
		50th - 74th percentile	-.3221	.25975	.604	-1.0076	.3633
		75th - 99th percentile	-1.5689*	.26521	.000	-2.2687	-.8690
	26th - 49th percentile	1st - 25th percentile	.2231	.25488	.818	-.4495	.8957
		50th - 74th percentile	-.0990	.27880	.984	-.8348	.6367
		75th - 99th percentile	-1.3458*	.28389	.000	-2.0949	-.5966
	50th - 74th percentile	1st - 25th percentile	.3221	.25975	.604	-.3633	1.0076
		26th - 49th percentile	.0990	.27880	.984	-.6367	.8348
		75th - 99th percentile	-1.2467*	.28827	.000	-2.0075	-.4860
	75th - 99th percentile	1st - 25th percentile	1.5689*	.26521	.000	.8690	2.2687
		26th - 49th percentile	1.3458*	.28389	.000	.5966	2.0949
		50th - 74th percentile	1.2467*	.28827	.000	.4860	2.0075
Zscore sentence Complexity 36m	1st - 25th percentile	26th - 49th percentile	-.2490	.27676	.805	-.9794	.4813
		50th - 74th percentile	-.4475	.28205	.393	-1.1918	.2968
		75th - 99th percentile	-.9282*	.28797	.011	-1.6882	-.1683
	26th - 49th percentile	1st - 25th percentile	.2490	.27676	.805	-.4813	.9794
		50th - 74th percentile	-.1985	.30273	.913	-.9974	.6004
		75th - 99th percentile	-.6792	.30826	.134	-1.4926	.1343
	50th - 74th percentile	1st - 25th percentile	.4475	.28205	.393	-.2968	1.1918
		26th - 49th percentile	.1985	.30273	.913	-.6004	.9974
		75th - 99th percentile	-.4807	.31302	.423	-1.3067	.3454
	75th - 99th percentile	1st - 25th percentile	.9282*	.28797	.011	.1683	1.6882
		26th - 49th percentile	.6792	.30826	.134	-.1343	1.4926
		50th - 74th percentile	.4807	.31302	.423	-.3454	1.3067

Based on observed means.

The error term is Mean Square(Error) = .710.

*. The mean difference is significant at the .05 level.

Gesture levels at 12m

Between-Subjects Factors

		Value Label	N
Gesture levels at 12m	1.00	1st - 25th percentile	11
	2.00	26th - 49th percentile	15
	3.00	50th - 74th percentile	17
	4.00	75th - 99th percentile	24

Descriptive Statistics

	Gesture levels at 12m	Mean	Std. Deviation	N
Zscore comprehension 12m	1st - 25th percentile	-.6070	.98063	11
	26th - 49th percentile	-.3835	.94579	15
	50th - 74th percentile	.1573	.69760	17
	75th - 99th percentile	.5615	1.11600	24
	Total	.0555	1.04889	67
Zscore production 12m	1st - 25th percentile	-.2498	.94307	11
	26th - 49th percentile	-.0700	.83266	15
	50th - 74th percentile	.0779	.85815	17
	75th - 99th percentile	.2949	1.23882	24
	Total	.0687	1.01738	67
Zscore comprehension 18m	1st - 25th percentile	-.7225	.54294	11
	26th - 49th percentile	-.0408	1.02098	15
	50th - 74th percentile	-.1627	1.02539	17
	75th - 99th percentile	.4719	.93425	24
	Total	.0000	1.00000	67
Zscore production 18m	1st - 25th percentile	-.4200	.51312	11
	26th - 49th percentile	-.0674	1.21019	15
	50th - 74th percentile	-.0082	.77142	17
	75th - 99th percentile	.3984	1.21629	24
	Total	.0566	1.04654	67
Zscore gestures 18m	1st - 25th percentile	-.5061	1.01809	11
	26th - 49th percentile	-.3834	.93574	15
	50th - 74th percentile	-.2274	.91723	17
	75th - 99th percentile	.6327	.79145	24
	Total	.0000	1.00000	67
Zscore Auditory Comprehension 18m	1st - 25th percentile	-.6074	.69887	11
	26th - 49th percentile	.0065	.95535	15
	50th - 74th percentile	-.0627	1.20467	17
	75th - 99th percentile	.3188	.89817	24
	Total	.0000	1.00000	67
Zscore Expressive Communication 18m	1st - 25th percentile	-.4652	.64158	11
	26th - 49th percentile	-.0978	.99453	15
	50th - 74th percentile	.0950	.97212	17
	75th - 99th percentile	.3329	1.00054	24
	Total	.0451	.96382	67

Zscore comprehension 24m	1st - 25th percentile	-.7785	.60183	11
	26th - 49th percentile	.0703	1.01446	15
	50th - 74th percentile	-.0314	.95753	17
	75th - 99th percentile	.3351	1.01826	24
	Total	.0000	1.00000	67
Zscore production 24m	1st - 25th percentile	-.4414	.79559	11
	26th - 49th percentile	.1061	.89742	15
	50th - 74th percentile	-.0141	1.18434	17
	75th - 99th percentile	.3010	1.10946	24
	Total	.0555	1.04911	67
Zscore sentence Complexity 24m	1st - 25th percentile	-.4707	.54812	11
	26th - 49th percentile	.0294	1.29374	15
	50th - 74th percentile	.1776	.98930	17
	75th - 99th percentile	.2965	.89078	24
	Total	.0806	.99246	67
Zscore Auditory Comprehension 24m	1st - 25th percentile	-.4682	.92588	11
	26th - 49th percentile	-.2356	1.07032	15
	50th - 74th percentile	.1470	1.07742	17
	75th - 99th percentile	.2577	.87507	24
	Total	.0000	1.00000	67
Zscore Expressive Communication 24m	1st - 25th percentile	-.0835	.82473	11
	26th - 49th percentile	-.2174	.86905	15
	50th - 74th percentile	.0749	1.17297	17
	75th - 99th percentile	.1211	1.05043	24
	Total	.0000	1.00000	67
Zscore ASQ 36m	1st - 25th percentile	-.3970	.85695	11
	26th - 49th percentile	-.0173	.93387	15
	50th - 74th percentile	.2487	.60015	17
	75th - 99th percentile	.0016	.97413	24
	Total	-.0054	.86977	67
Zscore production 36m	1st - 25th percentile	-.4356	.72837	11
	26th - 49th percentile	.3811	.85459	15
	50th - 74th percentile	-.1977	1.02325	17
	75th - 99th percentile	.1697	1.00211	24
	Total	.0244	.96107	67
Zscore sentence Complexity 36m	1st - 25th percentile	-.4540	.59625	11
	26th - 49th percentile	-.1175	.89670	15
	50th - 74th percentile	.2794	.89089	17
	75th - 99th percentile	.0902	.95259	24
	Total	.0023	.89086	67

Box's Test of Equality of Covariance Matrices[a]

Box's M	304.392
F	1.406
df1	120
df2	3702.400
Sig.	.003

Tests the null hypothesis that the observed covariance matrices of the dependent variables are equal across groups.

a. Design: Intercept + gesture levels at 12m

Multivariate Tests[a]

Effect		Value	F	Hypothesis df	Error df	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power[d]
Intercept	Pillai's Trace	.059	.206[b]	15.000	49.000	.999	.059	3.089	.121
	Wilks' Lambda	.941	.206[b]	15.000	49.000	.999	.059	3.089	.121
	Hotelling's Trace	.063	.206[b]	15.000	49.000	.999	.059	3.089	.121
	Roy's Largest Root	.063	.206[b]	15.000	49.000	.999	.059	3.089	.121
gesture levels at 12m	Pillai's Trace	.849	1.342	45.000	153.000	.097	.283	60.377	.985
	Wilks' Lambda	.359	1.340	45.000	146.347	.100	.289	59.602	.983
	Hotelling's Trace	1.261	1.336	45.000	143.000	.103	.296	60.121	.983
	Roy's Largest Root	.684	2.326[c]	15.000	51.000	.013	.406	34.896	.950

a. Design: Intercept + gesture levels at 12m

b. Exact statistic

c. The statistic is an upper bound on F that yields a lower bound on the significance level.

d. Computed using alpha = .05

Levene's Test of Equality of Error Variances[a]

	F	df1	df2	Sig.
Zscore comprehension 12m	.909	3	63	.442
Zscore production 12m	1.331	3	63	.272
Zscore comprehension 18m	1.942	3	63	.132
Zscore production 18m	1.981	3	63	.126
Zscore gestures 18m	.123	3	63	.946
Zscore Auditory Comprehension 18m	1.261	3	63	.295
Zscore Expressive Communication 18m	.912	3	63	.440
Zscore comprehension 24m	1.401	3	63	.251
Zscore production 24m	.681	3	63	.567
Zscore sentence Complexity 24m	1.550	3	63	.210
Zscore Auditory Comprehension 24m	.123	3	63	.946
Zscore Expressive Communication 24m	.360	3	63	.782
Zscore ASQ 36m	1.288	3	63	.286
Zscore production 36m	.601	3	63	.617
Zscore sentence Complexity 36m	1.526	3	63	.216

Tests the null hypothesis that the error variance of the dependent variable is equal across groups.

a. Design: Intercept + gesture levels at 12m

Tests of Between-Subjects Effects

Source	Dependent Variable	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power[p]
Corrected Model	Zscore comprehension 12m	14.040[a]	3	4.680	5.034	.003	.193	15.102	.900
	Zscore production 12m	2.634[b]	3	.878	.842	.476	.039	2.526	.223
	Zscore comprehension 18m	11.561[c]	3	3.854	4.460	.007	.175	13.379	.858

	Zscore production 18m	5.604[d]	3	1.868	1.765	.163	.078	5.294	.439
	Zscore gestures 18m	15.508[e]	3	5.169	6.450	.001	.235	19.350	.961
	Zscore Auditory Comprehension 18m	6.564[f]	3	2.188	2.319	.084	.099	6.958	.558
	Zscore Expressive Communication 18m	5.202[g]	3	1.734	1.947	.131	.085	5.840	.479
	Zscore comprehension 24m	9.453[h]	3	3.151	3.511	.020	.143	10.532	.756
	Zscore production 24m	4.283[i]	3	1.428	1.316	.277	.059	3.948	.334
	Zscore sentence Complexity 24m	4.661[j]	3	1.554	1.622	.193	.072	4.866	.406
	Zscore Auditory Comprehension 24m	5.204[k]	3	1.735	1.798	.157	.079	5.393	.446
	Zscore Expressive Communication 24m	1.233[l]	3	.411	.400	.754	.019	1.199	.125
	Zscore ASQ 36m	2.787[m]	3	.929	1.242	.302	.056	3.725	.317
	Zscore production 36m	5.581[n]	3	1.860	2.116	.107	.092	6.349	.516
	Zscore sentence Complexity 36m	3.997[o]	3	1.332	1.735	.169	.076	5.205	.432
Intercept	Zscore comprehension 12m	.286	1	.286	.308	.581	.005	.308	.085
	Zscore production 12m	.011	1	.011	.010	.919	.000	.010	.051
	Zscore comprehension 18m	.799	1	.799	.925	.340	.014	.925	.157
	Zscore production 18m	.037	1	.037	.035	.853	.001	.035	.054
	Zscore gestures 18m	.909	1	.909	1.134	.291	.018	1.134	.182
	Zscore Auditory Comprehension 18m	.461	1	.461	.488	.487	.008	.488	.106
	Zscore Expressive Communication 18m	.071	1	.071	.079	.779	.001	.079	.059
	Zscore comprehension 24m	.634	1	.634	.706	.404	.011	.706	.131
	Zscore production 24m	.009	1	.009	.008	.927	.000	.008	.051
	Zscore sentence Complexity 24m	.004	1	.004	.004	.948	.000	.004	.050
	Zscore Auditory Comprehension 24m	.347	1	.347	.359	.551	.006	.359	.091

	Zscore Expressive Communication 24m	.043	1	.043	.041	.839	.001	.041	.055
	Zscore ASQ 36m	.104	1	.104	.139	.710	.002	.139	.066
	Zscore production 36m	.026	1	.026	.030	.863	.000	.030	.053
	Zscore sentence Complexity 36m	.158	1	.158	.206	.652	.003	.206	.073
gesture levels at 12m	Zscore comprehension 12m	14.040	3	4.680	5.034	.003	.193	15.102	.900
	Zscore production 12m	2.634	3	.878	.842	.476	.039	2.526	.223
	Zscore comprehension 18m	11.561	3	3.854	4.460	.007	.175	13.379	.858
	Zscore production 18m	5.604	3	1.868	1.765	.163	.078	5.294	.439
	Zscore gestures 18m	15.508	3	5.169	6.450	.001	.235	19.350	.961
	Zscore Auditory Comprehension 18m	6.564	3	2.188	2.319	.084	.099	6.958	.558
	Zscore Expressive Communication 18m	5.202	3	1.734	1.947	.131	.085	5.840	.479
	Zscore comprehension 24m	9.453	3	3.151	3.511	.020	.143	10.532	.756
	Zscore production 24m	4.283	3	1.428	1.316	.277	.059	3.948	.334
	Zscore sentence Complexity 24m	4.661	3	1.554	1.622	.193	.072	4.866	.406
	Zscore Auditory Comprehension 24m	5.204	3	1.735	1.798	.157	.079	5.393	.446
	Zscore Expressive Communication 24m	1.233	3	.411	.400	.754	.019	1.199	.125
	Zscore ASQ 36m	2.787	3	.929	1.242	.302	.056	3.725	.317
	Zscore production 36m	5.581	3	1.860	2.116	.107	.092	6.349	.516
	Zscore sentence Complexity 36m	3.997	3	1.332	1.735	.169	.076	5.205	.432
Error	Zscore comprehension 12m	58.571	63	.930					
	Zscore production 12m	65.680	63	1.043					
	Zscore comprehension 18m	54.439	63	.864					
	Zscore production 18m	66.683	63	1.058					
	Zscore gestures 18m	50.492	63	.801					

	Zscore Auditory Comprehension 18m	59.436	63	.943						
	Zscore Expressive Communication 18m	56.109	63	.891						
	Zscore comprehension 24m	56.547	63	.898						
	Zscore production 24m	68.358	63	1.085						
	Zscore sentence Complexity 24m	60.347	63	.958						
	Zscore Auditory Comprehension 24m	60.796	63	.965						
	Zscore Expressive Communication 24m	64.767	63	1.028						
	Zscore ASQ 36m	47.141	63	.748						
	Zscore production 36m	55.379	63	.879						
	Zscore sentence Complexity 36m	48.382	63	.768						
Total	Zscore comprehension 12m	72.818	67							
	Zscore production 12m	68.631	67							
	Zscore comprehension 18m	66.000	67							
	Zscore production 18m	72.501	67							
	Zscore gestures 18m	66.000	67							
	Zscore Auditory Comprehension 18m	66.000	67							
	Zscore Expressive Communication 18m	61.446	67							
	Zscore comprehension 24m	66.000	67							
	Zscore production 24m	72.848	67							
	Zscore sentence Complexity 24m	65.443	67							
	Zscore Auditory Comprehension 24m	66.000	67							
	Zscore Expressive Communication 24m	66.000	67							
	Zscore ASQ 36m	49.931	67							

	Zscore production 36m	61.001	67						
	Zscore sentence Complexity 36m	52.380	67						
Corrected Total	Zscore comprehension 12m	72.612	66						
	Zscore production 12m	68.314	66						
	Zscore comprehension 18m	66.000	66						
	Zscore production 18m	72.287	66						
	Zscore gestures 18m	66.000	66						
	Zscore Auditory Comprehension 18m	66.000	66						
	Zscore Expressive Communication 18m	61.310	66						
	Zscore comprehension 24m	66.000	66						
	Zscore production 24m	72.641	66						
	Zscore sentence Complexity 24m	65.008	66						
	Zscore Auditory Comprehension 24m	66.000	66						
	Zscore Expressive Communication 24m	66.000	66						
	Zscore ASQ 36m	49.929	66						
	Zscore production 36m	60.961	66						
	Zscore sentence Complexity 36m	52.379	66						

a. R Squared = .193 (Adjusted R Squared = .155)

b. R Squared = .039 (Adjusted R Squared = -.007)

c. R Squared = .175 (Adjusted R Squared = .136)

d. R Squared = .078 (Adjusted R Squared = .034)

e. R Squared = .235 (Adjusted R Squared = .199)

f. R Squared = .099 (Adjusted R Squared = .057)

g. R Squared = .085 (Adjusted R Squared = .041)

h. R Squared = .143 (Adjusted R Squared = .102)

i. R Squared = .059 (Adjusted R Squared = .014)

j. R Squared = .072 (Adjusted R Squared = .028)

k. R Squared = .079 (Adjusted R Squared = .035)

l. R Squared = .019 (Adjusted R Squared = -.028)

m. R Squared = .056 (Adjusted R Squared = .011)

n. R Squared = .092 (Adjusted R Squared = .048)

o. R Squared = .076 (Adjusted R Squared = .032)

p. Computed using alpha = .05

Gesture levels at 12m

Dependent Variable	Gesture levels at 12m	Mean	Std. Error	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
Zscore comprehension 12m	1st - 25th percentile	-.607	.291	-1.188	-.026
	26th - 49th percentile	-.384	.249	-.881	.114
	50th - 74th percentile	.157	.234	-.310	.625
	75th - 99th percentile	.562	.197	.168	.955
Zscore production 12m	1st - 25th percentile	-.250	.308	-.865	.365
	26th - 49th percentile	-.070	.264	-.597	.457
	50th - 74th percentile	.078	.248	-.417	.573
	75th - 99th percentile	.295	.208	-.122	.711
Zscore comprehension 18m	1st - 25th percentile	-.722	.280	-1.283	-.162
	26th - 49th percentile	-.041	.240	-.520	.439
	50th - 74th percentile	-.163	.225	-.613	.288
	75th - 99th percentile	.472	.190	.093	.851
Zscore production 18m	1st - 25th percentile	-.420	.310	-1.040	.200
	26th - 49th percentile	-.067	.266	-.598	.463
	50th - 74th percentile	-.008	.250	-.507	.490
	75th - 99th percentile	.398	.210	-.021	.818
Zscore gestures 18m	1st - 25th percentile	-.506	.270	-1.046	.033
	26th - 49th percentile	-.383	.231	-.845	.078
	50th - 74th percentile	-.227	.217	-.661	.207
	75th - 99th percentile	.633	.183	.267	.998
Zscore Auditory Comprehension 18m	1st - 25th percentile	-.607	.293	-1.193	-.022
	26th - 49th percentile	.007	.251	-.495	.508
	50th - 74th percentile	-.063	.236	-.534	.408
	75th - 99th percentile	.319	.198	-.077	.715
Zscore Expressive Communication 18m	1st - 25th percentile	-.465	.285	-1.034	.103
	26th - 49th percentile	-.098	.244	-.585	.389
	50th - 74th percentile	.095	.229	-.362	.552
	75th - 99th percentile	.333	.193	-.052	.718
Zscore comprehension 24m	1st - 25th percentile	-.779	.286	-1.349	-.208
	26th - 49th percentile	.070	.245	-.419	.559
	50th - 74th percentile	-.031	.230	-.491	.428
	75th - 99th percentile	.335	.193	-.051	.722

Zscore production 24m	1st - 25th percentile	-.441	.314	-1.069	.186
	26th - 49th percentile	.106	.269	-.431	.644
	50th - 74th percentile	-.014	.253	-.519	.491
	75th - 99th percentile	.301	.213	-.124	.726
Zscore sentence Complexity 24m	1st - 25th percentile	-.471	.295	-1.060	.119
	26th - 49th percentile	.029	.253	-.476	.534
	50th - 74th percentile	.178	.237	-.297	.652
	75th - 99th percentile	.296	.200	-.103	.696
Zscore Auditory Comprehension 24m	1st - 25th percentile	-.468	.296	-1.060	.124
	26th - 49th percentile	-.236	.254	-.742	.271
	50th - 74th percentile	.147	.238	-.329	.623
	75th - 99th percentile	.258	.201	-.143	.658
Zscore Expressive Communication 24m	1st - 25th percentile	-.084	.306	-.694	.527
	26th - 49th percentile	-.217	.262	-.741	.306
	50th - 74th percentile	.075	.246	-.417	.566
	75th - 99th percentile	.121	.207	-.293	.535
Zscore ASQ 36m	1st - 25th percentile	-.397	.261	-.918	.124
	26th - 49th percentile	-.017	.223	-.464	.429
	50th - 74th percentile	.249	.210	-.171	.668
	75th - 99th percentile	.002	.177	-.351	.354
Zscore production 36m	1st - 25th percentile	-.436	.283	-1.000	.129
	26th - 49th percentile	.381	.242	-.103	.865
	50th - 74th percentile	-.198	.227	-.652	.257
	75th - 99th percentile	.170	.191	-.213	.552
Zscore sentence Complexity 36m	1st - 25th percentile	-.454	.264	-.982	.074
	26th - 49th percentile	-.118	.226	-.570	.335
	50th - 74th percentile	.279	.213	-.145	.704
	75th - 99th percentile	.090	.179	-.267	.448

Multiple Comparisons

Tukey HSD

Dependent Variable	(I) Gesture levels at 12m	(J) Gesture levels at 12m	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
Zscore comprehension 12m	1st - 25th percentile	26th - 49th percentile	-.2235	.38275	.937	-1.2335	.7866
		50th - 74th percentile	-.7642	.37310	.182	-1.7488	.2204
		75th - 99th percentile	-1.1685*	.35108	.008	-2.0950	-.2420
	26th - 49th percentile	1st - 25th percentile	.2235	.38275	.937	-.7866	1.2335

		50th - 74th percentile	-.5408	.34157	.395	-1.4421	.3606
		75th - 99th percentile	-.9450*	.31736	.021	-1.7825	-.1075
	50th - 74th percentile	1st - 25th percentile	.7642	.37310	.182	-.2204	1.7488
		26th - 49th percentile	.5408	.34157	.395	-.3606	1.4421
		75th - 99th percentile	-.4043	.30566	.552	-1.2109	.4023
	75th - 99th percentile	1st - 25th percentile	1.1685*	.35108	.008	.2420	2.0950
		26th - 49th percentile	.9450*	.31736	.021	.1075	1.7825
		50th - 74th percentile	.4043	.30566	.552	-.4023	1.2109
Zscore production 12m	1st - 25th percentile	26th - 49th percentile	-.1798	.40531	.971	-1.2495	.8898
		50th - 74th percentile	-.3278	.39510	.840	-1.3704	.7149
		75th - 99th percentile	-.5447	.37177	.464	-1.5258	.4364
	26th - 49th percentile	1st - 25th percentile	.1798	.40531	.971	-.8898	1.2495
		50th - 74th percentile	-.1479	.36170	.977	-1.1024	.8066
		75th - 99th percentile	-.3649	.33607	.699	-1.2517	.5220
	50th - 74th percentile	1st - 25th percentile	.3278	.39510	.840	-.7149	1.3704
		26th - 49th percentile	.1479	.36170	.977	-.8066	1.1024
		75th - 99th percentile	-.2169	.32368	.908	-1.0711	.6372
	75th - 99th percentile	1st - 25th percentile	.5447	.37177	.464	-.4364	1.5258
		26th - 49th percentile	.3649	.33607	.699	-.5220	1.2517
		50th - 74th percentile	.2169	.32368	.908	-.6372	1.0711
Zscore comprehension 18m	1st - 25th percentile	26th - 49th percentile	-.6816	.36900	.261	-1.6554	.2921
		50th - 74th percentile	-.5597	.35970	.411	-1.5090	.3895
		75th - 99th percentile	-1.1944*	.33847	.004	-2.0876	-.3012
	26th - 49th percentile	1st - 25th percentile	.6816	.36900	.261	-.2921	1.6554
		50th - 74th percentile	.1219	.32930	.983	-.7471	.9909
		75th - 99th percentile	-.5127	.30596	.345	-1.3201	.2947
	50th - 74th percentile	1st - 25th percentile	.5597	.35970	.411	-.3895	1.5090
		26th - 49th percentile	-.1219	.32930	.983	-.9909	.7471
		75th - 99th percentile	-.6346	.29468	.148	-1.4123	.1430
	75th - 99th percentile	1st - 25th percentile	1.1944*	.33847	.004	.3012	2.0876
		26th - 49th percentile	.5127	.30596	.345	-.2947	1.3201
		50th - 74th percentile	.6346	.29468	.148	-.1430	1.4123
Zscore production 18m	1st - 25th percentile	26th - 49th percentile	-.3526	.40840	.824	-1.4303	.7252
		50th - 74th percentile	-.4118	.39810	.730	-1.4623	.6388
		75th - 99th percentile	-.8183	.37460	.139	-1.8069	.1702
	26th - 49th percentile	1st - 25th percentile	.3526	.40840	.824	-.7252	1.4303
		50th - 74th percentile	-.0592	.36445	.998	-1.0210	.9026
		75th - 99th percentile	-.4658	.33862	.519	-1.3594	.4279
	50th - 74th percentile	1st - 25th percentile	.4118	.39810	.730	-.6388	1.4623
		26th - 49th percentile	.0592	.36445	.998	-.9026	1.0210
		75th - 99th percentile	-.4066	.32614	.600	-1.2672	.4541
	75th - 99th percentile	1st - 25th percentile	.8183	.37460	.139	-.1702	1.8069

		26th - 49th percentile	.4658	.33862	.519	-.4279	1.3594
		50th - 74th percentile	.4066	.32614	.600	-.4541	1.2672
Zscore gestures 18m	1st - 25th percentile	26th - 49th percentile	-.1227	.35537	.986	-1.0605	.8151
		50th - 74th percentile	-.2787	.34642	.852	-1.1929	.6354
		75th - 99th percentile	-1.1388*	.32597	.005	-1.9990	-.2786
	26th - 49th percentile	1st - 25th percentile	.1227	.35537	.986	-.8151	1.0605
		50th - 74th percentile	-.1560	.31714	.961	-.9929	.6809
		75th - 99th percentile	-1.0161*	.29466	.005	-1.7937	-.2385
	50th - 74th percentile	1st - 25th percentile	.2787	.34642	.852	-.6354	1.1929
		26th - 49th percentile	.1560	.31714	.961	-.6809	.9929
		75th - 99th percentile	-.8600*	.28379	.018	-1.6090	-.1111
	75th - 99th percentile	1st - 25th percentile	1.1388*	.32597	.005	.2786	1.9990
		26th - 49th percentile	1.0161*	.29466	.005	.2385	1.7937
		50th - 74th percentile	.8600*	.28379	.018	.1111	1.6090
Zscore Auditory Comprehension 18m	1st - 25th percentile	26th - 49th percentile	-.6139	.38557	.390	-1.6314	.4036
		50th - 74th percentile	-.5446	.37585	.474	-1.5365	.4472
		75th - 99th percentile	-.9262	.35366	.053	-1.8594	.0071
	26th - 49th percentile	1st - 25th percentile	.6139	.38557	.390	-.4036	1.6314
		50th - 74th percentile	.0692	.34408	.997	-.8388	.9773
		75th - 99th percentile	-.3123	.31969	.763	-1.1559	.5314
	50th - 74th percentile	1st - 25th percentile	.5446	.37585	.474	-.4472	1.5365
		26th - 49th percentile	-.0692	.34408	.997	-.9773	.8388
		75th - 99th percentile	-.3815	.30790	.605	-1.1941	.4310
	75th - 99th percentile	1st - 25th percentile	.9262	.35366	.053	-.0071	1.8594
		26th - 49th percentile	.3123	.31969	.763	-.5314	1.1559
		50th - 74th percentile	.3815	.30790	.605	-.4310	1.1941
Zscore Expressive Communication 18m	1st - 25th percentile	26th - 49th percentile	-.3675	.37462	.761	-1.3561	.6211
		50th - 74th percentile	-.5602	.36518	.424	-1.5239	.4035
		75th - 99th percentile	-.7982	.34362	.104	-1.7050	.1086
	26th - 49th percentile	1st - 25th percentile	.3675	.37462	.761	-.6211	1.3561
		50th - 74th percentile	-.1927	.33431	.939	-1.0750	.6895
		75th - 99th percentile	-.4307	.31062	.512	-1.2504	.3890
	50th - 74th percentile	1st - 25th percentile	.5602	.36518	.424	-.4035	1.5239
		26th - 49th percentile	.1927	.33431	.939	-.6895	1.0750
		75th - 99th percentile	-.2380	.29916	.856	-1.0274	.5515
	75th - 99th percentile	1st - 25th percentile	.7982	.34362	.104	-.1086	1.7050
		26th - 49th percentile	.4307	.31062	.512	-.3890	1.2504
		50th - 74th percentile	.2380	.29916	.856	-.5515	1.0274
Zscore comprehension 24m	1st - 25th percentile	26th - 49th percentile	-.8488	.37608	.119	-1.8413	.1436
		50th - 74th percentile	-.7472	.36660	.185	-1.7146	.2203
		75th - 99th percentile	-1.1136*	.34496	.010	-2.0239	-.2033
	26th - 49th percentile	1st - 25th percentile	.8488	.37608	.119	-.1436	1.8413

		50th - 74th percentile	.1016	.33561	.990	-.7840	.9873
		75th - 99th percentile	-.2648	.31183	.831	-1.0877	.5581
	50th - 74th percentile	1st - 25th percentile	.7472	.36660	.185	-.2203	1.7146
		26th - 49th percentile	-.1016	.33561	.990	-.9873	.7840
		75th - 99th percentile	-.3664	.30033	.617	-1.1590	.4261
	75th - 99th percentile	1st - 25th percentile	1.1136*	.34496	.010	.2033	2.0239
		26th - 49th percentile	.2648	.31183	.831	-.5581	1.0877
		50th - 74th percentile	.3664	.30033	.617	-.4261	1.1590
Zscore production 24m	1st - 25th percentile	26th - 49th percentile	-.5475	.41349	.551	-1.6387	.5436
		50th - 74th percentile	-.4274	.40307	.715	-1.4910	.6363
		75th - 99th percentile	-.7424	.37928	.215	-1.7433	.2585
	26th - 49th percentile	1st - 25th percentile	.5475	.41349	.551	-.5436	1.6387
		50th - 74th percentile	.1202	.36900	.988	-.8536	1.0940
		75th - 99th percentile	-.1949	.34285	.941	-1.0996	.7099
	50th - 74th percentile	1st - 25th percentile	.4274	.40307	.715	-.6363	1.4910
		26th - 49th percentile	-.1202	.36900	.988	-1.0940	.8536
		75th - 99th percentile	-.3150	.33021	.776	-1.1864	.5564
	75th - 99th percentile	1st - 25th percentile	.7424	.37928	.215	-.2585	1.7433
		26th - 49th percentile	.1949	.34285	.941	-.7099	1.0996
		50th - 74th percentile	.3150	.33021	.776	-.5564	1.1864
Zscore sentence Complexity 24m	1st - 25th percentile	26th - 49th percentile	-.5001	.38851	.574	-1.5254	.5251
		50th - 74th percentile	-.6483	.37872	.326	-1.6477	.3511
		75th - 99th percentile	-.7672	.35636	.148	-1.7076	.1732
	26th - 49th percentile	1st - 25th percentile	.5001	.38851	.574	-.5251	1.5254
		50th - 74th percentile	-.1481	.34671	.974	-1.0631	.7668
		75th - 99th percentile	-.2671	.32214	.840	-1.1172	.5830
	50th - 74th percentile	1st - 25th percentile	.6483	.37872	.326	-.3511	1.6477
		26th - 49th percentile	.1481	.34671	.974	-.7668	1.0631
		75th - 99th percentile	-.1189	.31026	.981	-.9377	.6998
	75th - 99th percentile	1st - 25th percentile	.7672	.35636	.148	-.1732	1.7076
		26th - 49th percentile	.2671	.32214	.840	-.5830	1.1172
		50th - 74th percentile	.1189	.31026	.981	-.6998	.9377
Zscore Auditory Comprehension 24m	1st - 25th percentile	26th - 49th percentile	-.2326	.38995	.933	-1.2616	.7965
		50th - 74th percentile	-.6152	.38012	.376	-1.6183	.3879
		75th - 99th percentile	-.7258	.35768	.188	-1.6697	.2181
	26th - 49th percentile	1st - 25th percentile	.2326	.38995	.933	-.7965	1.2616
		50th - 74th percentile	-.3826	.34799	.691	-1.3009	.5357
		75th - 99th percentile	-.4932	.32333	.429	-1.3465	.3600
	50th - 74th percentile	1st - 25th percentile	.6152	.38012	.376	-.3879	1.6183
		26th - 49th percentile	.3826	.34799	.691	-.5357	1.3009
		75th - 99th percentile	-.1106	.31141	.984	-.9324	.7112
	75th - 99th percentile	1st - 25th percentile	.7258	.35768	.188	-.2181	1.6697

		26th - 49th percentile	.4932	.32333	.429	-.3600	1.3465
		50th - 74th percentile	.1106	.31141	.984	-.7112	.9324
Zscore Expressive Communication 24m	1st - 25th percentile	26th - 49th percentile	.1339	.40249	.987	-.9283	1.1960
		50th - 74th percentile	-.1584	.39234	.978	-1.1938	.8770
		75th - 99th percentile	-.2046	.36918	.945	-1.1789	.7697
	26th - 49th percentile	1st - 25th percentile	-.1339	.40249	.987	-1.1960	.9283
		50th - 74th percentile	-.2923	.35918	.848	-1.2401	.6556
		75th - 99th percentile	-.3385	.33372	.742	-1.2191	.5422
	50th - 74th percentile	1st - 25th percentile	.1584	.39234	.978	-.8770	1.1938
		26th - 49th percentile	.2923	.35918	.848	-.6556	1.2401
		75th - 99th percentile	-.0462	.32142	.999	-.8944	.8020
	75th - 99th percentile	1st - 25th percentile	.2046	.36918	.945	-.7697	1.1789
		26th - 49th percentile	.3385	.33372	.742	-.5422	1.2191
		50th - 74th percentile	.0462	.32142	.999	-.8020	.8944
Zscore ASQ 36m	1st - 25th percentile	26th - 49th percentile	-.3797	.34338	.687	-1.2859	.5265
		50th - 74th percentile	-.6457	.33473	.227	-1.5290	.2377
		75th - 99th percentile	-.3986	.31497	.588	-1.2298	.4326
	26th - 49th percentile	1st - 25th percentile	.3797	.34338	.687	-.5265	1.2859
		50th - 74th percentile	-.2660	.30643	.821	-1.0746	.5427
		75th - 99th percentile	-.0189	.28472	1.000	-.7702	.7325
	50th - 74th percentile	1st - 25th percentile	.6457	.33473	.227	-.2377	1.5290
		26th - 49th percentile	.2660	.30643	.821	-.5427	1.0746
		75th - 99th percentile	.2471	.27422	.804	-.4766	.9707
	75th - 99th percentile	1st - 25th percentile	.3986	.31497	.588	-.4326	1.2298
		26th - 49th percentile	.0189	.28472	1.000	-.7325	.7702
		50th - 74th percentile	-.2471	.27422	.804	-.9707	.4766
Zscore production 36m	1st - 25th percentile	26th - 49th percentile	-.8167	.37218	.136	-1.7989	.1655
		50th - 74th percentile	-.2379	.36280	.913	-1.1953	.7195
		75th - 99th percentile	-.6053	.34138	.296	-1.5061	.2956
	26th - 49th percentile	1st - 25th percentile	.8167	.37218	.136	-.1655	1.7989
		50th - 74th percentile	.5788	.33213	.311	-.2977	1.4553
		75th - 99th percentile	.2114	.30859	.902	-.6029	1.0258
	50th - 74th percentile	1st - 25th percentile	.2379	.36280	.913	-.7195	1.1953
		26th - 49th percentile	-.5788	.33213	.311	-1.4553	.2977
		75th - 99th percentile	-.3674	.29721	.607	-1.1517	.4169
	75th - 99th percentile	1st - 25th percentile	.6053	.34138	.296	-.2956	1.5061
		26th - 49th percentile	-.2114	.30859	.902	-1.0258	.6029
		50th - 74th percentile	.3674	.29721	.607	-.4169	1.1517
Zscore sentence Complexity 36m	1st - 25th percentile	26th - 49th percentile	-.3365	.34787	.768	-1.2545	.5815
		50th - 74th percentile	-.7335	.33910	.145	-1.6283	.1614

		75th - 99th percentile	-.5442	.31908	.329	-1.3863	.2978
	26th - 49th percentile	1st - 25th percentile	.3365	.34787	.768	-.5815	1.2545
		50th - 74th percentile	-.3970	.31044	.580	-1.2162	.4223
		75th - 99th percentile	-.2077	.28844	.889	-.9689	.5535
	50th - 74th percentile	1st - 25th percentile	.7335	.33910	.145	-.1614	1.6283
		26th - 49th percentile	.3970	.31044	.580	-.4223	1.2162
		75th - 99th percentile	.1892	.27780	.904	-.5439	.9223
	75th - 99th percentile	1st - 25th percentile	.5442	.31908	.329	-.2978	1.3863
		26th - 49th percentile	.2077	.28844	.889	-.5535	.9689
		50th - 74th percentile	-.1892	.27780	.904	-.9223	.5439

Based on observed means.

The error term is Mean Square(Error) = .768.

*. The mean difference is significant at the .05 level.

Gesture levels at 18m

Between-Subjects Factors

		Value Label	N
gesture levels at 18m	1.00	1st - 25th percentile	12
	2.00	26th - 49th percentile	18
	3.00	50th - 74th percentile	19
	4.00	75th - 99th percentile	18

Descriptive Statistics

	gesture levels at 18m	Mean	Std. Deviation	N
Zscore comprehension 18m	1st - 25th percentile	-.9333	.64042	12
	26th - 49th percentile	-.2030	1.02763	18
	50th - 74th percentile	.2201	.97542	19
	75th - 99th percentile	.5929	.68468	18
	Total	.0000	1.00000	67
Zscore production 18m	1st - 25th percentile	-.4660	1.12512	12
	26th - 49th percentile	-.2924	.75536	18
	50th - 74th percentile	.3077	1.01581	19
	75th - 99th percentile	.4888	1.08621	18
	Total	.0566	1.04654	67
Zscore Auditory Comprehension 18m	1st - 25th percentile	-.7501	.86855	12
	26th - 49th percentile	.0590	.96524	18
	50th - 74th percentile	.2009	.92371	19
	75th - 99th percentile	.2290	1.02840	18
	Total	.0000	1.00000	67
	1st - 25th percentile	-.2364	1.16684	12

Zscore Expressive Communication 18m	26th - 49th percentile	-.0449	.72536	18
	50th - 74th percentile	.1044	.91823	19
	75th - 99th percentile	.2600	1.09054	18
	Total	.0451	.96382	67
Zscore comprehension 24m	1st - 25th percentile	-.5648	.66253	12
	26th - 49th percentile	-.1838	1.11592	18
	50th - 74th percentile	.2310	.98481	19
	75th - 99th percentile	.3166	.94838	18
	Total	.0000	1.00000	67
Zscore production 24m	1st - 25th percentile	-.3873	.83280	12
	26th - 49th percentile	-.1424	.95036	18
	50th - 74th percentile	.2123	.99036	19
	75th - 99th percentile	.3832	1.24960	18
	Total	.0555	1.04911	67
Zscore sentence Complexity 24m	1st - 25th percentile	-.2440	.83824	12
	26th - 49th percentile	-.1230	.78016	18
	50th - 74th percentile	.3219	1.27247	19
	75th - 99th percentile	.2457	.91328	18
	Total	.0806	.99246	67
Zscore Auditory Comprehension 24m	1st - 25th percentile	-.5912	1.17964	12
	26th - 49th percentile	-.0277	.95689	18
	50th - 74th percentile	.0100	.76547	19
	75th - 99th percentile	.4113	1.00949	18
	Total	.0000	1.00000	67
Zscore Expressive Communication 24m	1st - 25th percentile	-.1057	.78250	12
	26th - 49th percentile	-.1522	.87647	18
	50th - 74th percentile	-.0194	1.12325	19
	75th - 99th percentile	.2432	1.13427	18
	Total	.0000	1.00000	67
Zscore ASQ 36m	1st - 25th percentile	-.3932	.92059	12
	26th - 49th percentile	.0899	.73302	18
	50th - 74th percentile	.0465	.73704	19
	75th - 99th percentile	.1031	1.06999	18
	Total	-.0054	.86977	67
Zscore production 36m	1st - 25th percentile	-.2908	.73308	12
	26th - 49th percentile	-.1267	.97943	18
	50th - 74th percentile	.2169	.88349	19
	75th - 99th percentile	.1825	1.13590	18
	Total	.0244	.96107	67
Zscore sentence Complexity 36m	1st - 25th percentile	-.0824	1.02272	12

	26th - 49th percentile	.0172	.96504	18
	50th - 74th percentile	-.0250	.77326	19
	75th - 99th percentile	.0728	.91023	18
	Total	.0023	.89086	67

Box's Test of Equality of Covariance Matrices[a]

Box's M	265.885
F	1.124
df1	156
df2	7168.420
Sig.	.141

Tests the null hypothesis that the observed covariance matrices of the dependent variables are equal across groups.

a. Design: Intercept + gesture levels at 18m

Multivariate Tests[a]

Effect		Value	F	Hypothesis df	Error df	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power[d]
Intercept	Pillai's Trace	.036	.162[b]	12.000	52.000	.999	.036	1.938	.099
	Wilks' Lambda	.964	.162[b]	12.000	52.000	.999	.036	1.938	.099
	Hotelling's Trace	.037	.162[b]	12.000	52.000	.999	.036	1.938	.099
	Roy's Largest Root	.037	.162[b]	12.000	52.000	.999	.036	1.938	.099
gesture levels at 18m	Pillai's Trace	.661	1.271	36.000	162.000	.160	.220	45.747	.959
	Wilks' Lambda	.450	1.329	36.000	154.367	.121	.233	47.012	.964
	Hotelling's Trace	.987	1.390	36.000	152.000	.089	.248	50.033	.975
	Roy's Largest Root	.705	3.173[c]	12.000	54.000	.002	.414	38.077	.983

a. Design: Intercept + gesture levels at 18m

b. Exact statistic

c. The statistic is an upper bound on F that yields a lower bound on the significance level.

d. Computed using alpha = .05

Levene's Test of Equality of Error Variances[a]

	F	df1	df2	Sig.
Zscore comprehension 18m	1.302	3	63	.282
Zscore production 18m	1.043	3	63	.380
Zscore Auditory Comprehension 18m	.153	3	63	.928
Zscore Expressive Communication 18m	2.908	3	63	.041
Zscore comprehension 24m	1.275	3	63	.291
Zscore production 24m	.457	3	63	.713
Zscore sentence Complexity 24m	1.117	3	63	.349
Zscore Auditory Comprehension 24m	1.126	3	63	.345
Zscore Expressive Communication 24m	.732	3	63	.537
Zscore ASQ 36m	1.583	3	63	.202
Zscore production 36m	1.135	3	63	.342
Zscore sentence Complexity 36m	.459	3	63	.712

Tests the null hypothesis that the error variance of the dependent variable is equal across groups.

a. Design: Intercept + gesture levels at 18m

Tests of Between-Subjects Effects

Source	Dependent Variable	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power[m]
Corrected Model	Zscore comprehension 18m	18.441[a]	3	6.147	8.143	.000	.279	24.428	.988
	Zscore production 18m	10.031[b]	3	3.344	3.384	.023	.139	10.151	.739
	Zscore Auditory Comprehension 18m	8.525[c]	3	2.842	3.115	.032	.129	9.345	.699
	Zscore Expressive Communication 18m	1.995[d]	3	.665	.706	.552	.033	2.119	.192
	Zscore comprehension 24m	7.254[e]	3	2.418	2.593	.060	.110	7.779	.611
	Zscore production 24m	5.458[f]	3	1.819	1.706	.175	.075	5.118	.425
	Zscore sentence Complexity 24m	3.608[g]	3	1.203	1.234	.305	.055	3.702	.315
	Zscore Auditory Comprehension 24m	7.256[h]	3	2.419	2.594	.060	.110	7.782	.611
	Zscore Expressive Communication 24m	1.623[i]	3	.541	.530	.664	.025	1.589	.152

	Zscore ASQ 36m	2.231[j]	3	.744	.982	.407	.045	2.947	.256
	Zscore production 36m	2.757[k]	3	.919	.995	.401	.045	2.984	.258
	Zscore sentence Complexity 36m	.194[l]	3	.065	.078	.972	.004	.234	.063
Intercept	Zscore comprehension 18m	.423	1	.423	.560	.457	.009	.560	.114
	Zscore production 18m	.006	1	.006	.006	.939	.000	.006	.051
	Zscore Auditory Comprehension 18m	.276	1	.276	.303	.584	.005	.303	.084
	Zscore Expressive Communication 18m	.028	1	.028	.030	.864	.000	.030	.053
	Zscore comprehension 24m	.164	1	.164	.176	.677	.003	.176	.070
	Zscore production 24m	.018	1	.018	.016	.898	.000	.016	.052
	Zscore sentence Complexity 24m	.163	1	.163	.167	.684	.003	.167	.069
	Zscore Auditory Comprehension 24m	.158	1	.158	.170	.682	.003	.170	.069
	Zscore Expressive Communication 24m	.005	1	.005	.005	.946	.000	.005	.051
	Zscore ASQ 36m	.096	1	.096	.126	.724	.002	.126	.064
	Zscore production 36m	.001	1	.001	.001	.970	.000	.001	.050
	Zscore sentence Complexity 36m	.001	1	.001	.001	.970	.000	.001	.050
gesture levels at 18m	Zscore comprehension 18m	18.441	3	6.147	8.143	.000	.279	24.428	.988
	Zscore production 18m	10.031	3	3.344	3.384	.023	.139	10.151	.739
	Zscore Auditory Comprehension 18m	8.525	3	2.842	3.115	.032	.129	9.345	.699
	Zscore Expressive Communication 18m	1.995	3	.665	.706	.552	.033	2.119	.192
	Zscore comprehension 24m	7.254	3	2.418	2.593	.060	.110	7.779	.611
	Zscore production 24m	5.458	3	1.819	1.706	.175	.075	5.118	.425
	Zscore sentence Complexity 24m	3.608	3	1.203	1.234	.305	.055	3.702	.315
	Zscore Auditory Comprehension 24m	7.256	3	2.419	2.594	.060	.110	7.782	.611
	Zscore Expressive Communication 24m	1.623	3	.541	.530	.664	.025	1.589	.152
	Zscore ASQ 36m	2.231	3	.744	.982	.407	.045	2.947	.256
	Zscore production 36m	2.757	3	.919	.995	.401	.045	2.984	.258
	Zscore sentence Complexity 36m	.194	3	.065	.078	.972	.004	.234	.063
Error	Zscore comprehension 18m	47.559	63	.755					

	Zscore production 18m	62.255	63	.988					
	Zscore Auditory Comprehension 18m	57.475	63	.912					
	Zscore Expressive Communication 18m	59.315	63	.942					
	Zscore comprehension 24m	58.746	63	.932					
	Zscore production 24m	67.183	63	1.066					
	Zscore sentence Complexity 24m	61.401	63	.975					
	Zscore Auditory Comprehension 24m	58.744	63	.932					
	Zscore Expressive Communication 24m	64.377	63	1.022					
	Zscore ASQ 36m	47.698	63	.757					
	Zscore production 36m	58.204	63	.924					
	Zscore sentence Complexity 36m	52.185	63	.828					
Total	Zscore comprehension 18m	66.000	67						
	Zscore production 18m	72.501	67						
	Zscore Auditory Comprehension 18m	66.000	67						
	Zscore Expressive Communication 18m	61.446	67						
	Zscore comprehension 24m	66.000	67						
	Zscore production 24m	72.848	67						
	Zscore sentence Complexity 24m	65.443	67						
	Zscore Auditory Comprehension 24m	66.000	67						
	Zscore Expressive Communication 24m	66.000	67						
	Zscore ASQ 36m	49.931	67						
	Zscore production 36m	61.001	67						
	Zscore sentence Complexity 36m	52.380	67						
Corrected Total	Zscore comprehension 18m	66.000	66						
	Zscore production 18m	72.287	66						
	Zscore Auditory Comprehension 18m	66.000	66						
	Zscore Expressive Communication 18m	61.310	66						

	Zscore comprehension 24m	66.000	66							
	Zscore production 24m	72.641	66							
	Zscore sentence Complexity 24m	65.008	66							
	Zscore Auditory Comprehension 24m	66.000	66							
	Zscore Expressive Communication 24m	66.000	66							
	Zscore ASQ 36m	49.929	66							
	Zscore production 36m	60.961	66							
	Zscore sentence Complexity 36m	52.379	66							

a. R Squared = .279 (Adjusted R Squared = .245)

b. R Squared = .139 (Adjusted R Squared = .098)

c. R Squared = .129 (Adjusted R Squared = .088)

d. R Squared = .033 (Adjusted R Squared = -.014)

e. R Squared = .110 (Adjusted R Squared = .068)

f. R Squared = .075 (Adjusted R Squared = .031)

g. R Squared = .055 (Adjusted R Squared = .011)

h. R Squared = .110 (Adjusted R Squared = .068)

i. R Squared = .025 (Adjusted R Squared = -.022)

j. R Squared = .045 (Adjusted R Squared = -.001)

k. R Squared = .045 (Adjusted R Squared = .000)

l. R Squared = .004 (Adjusted R Squared = -.044)

m. Computed using alpha = .05

gesture levels at 18m

Dependent Variable	gesture levels at 18m	Mean	Std. Error	95% Confidence Interval	
				Lower Bound	Upper Bound
Zscore comprehension 18m	1st - 25th percentile	-.933	.251	-1.434	-.432
	26th - 49th percentile	-.203	.205	-.612	.206
	50th - 74th percentile	.220	.199	-.178	.618
	75th - 99th percentile	.593	.205	.184	1.002
Zscore production 18m	1st - 25th percentile	-.466	.287	-1.039	.107
	26th - 49th percentile	-.292	.234	-.761	.176
	50th - 74th percentile	.308	.228	-.148	.763
	75th - 99th percentile	.489	.234	.021	.957
Zscore Auditory Comprehension 18m	1st - 25th percentile	-.750	.276	-1.301	-.199
	26th - 49th percentile	.059	.225	-.391	.509
	50th - 74th percentile	.201	.219	-.237	.639
	75th - 99th percentile	.229	.225	-.221	.679
	1st - 25th percentile	-.236	.280	-.796	.323

Zscore Expressive Communication 18m	26th - 49th percentile	-.045	.229	-.502	.412
	50th - 74th percentile	.104	.223	-.340	.549
	75th - 99th percentile	.260	.229	-.197	.717
Zscore comprehension 24m	1st - 25th percentile	-.565	.279	-1.122	-.008
	26th - 49th percentile	-.184	.228	-.639	.271
	50th - 74th percentile	.231	.222	-.212	.674
	75th - 99th percentile	.317	.228	-.138	.771
Zscore production 24m	1st - 25th percentile	-.387	.298	-.983	.208
	26th - 49th percentile	-.142	.243	-.629	.344
	50th - 74th percentile	.212	.237	-.261	.686
	75th - 99th percentile	.383	.243	-.103	.870
Zscore sentence Complexity 24m	1st - 25th percentile	-.244	.285	-.813	.326
	26th - 49th percentile	-.123	.233	-.588	.342
	50th - 74th percentile	.322	.226	-.131	.775
	75th - 99th percentile	.246	.233	-.219	.711
Zscore Auditory Comprehension 24m	1st - 25th percentile	-.591	.279	-1.148	-.034
	26th - 49th percentile	-.028	.228	-.483	.427
	50th - 74th percentile	.010	.222	-.433	.453
	75th - 99th percentile	.411	.228	-.043	.866
Zscore Expressive Communication 24m	1st - 25th percentile	-.106	.292	-.689	.477
	26th - 49th percentile	-.152	.238	-.628	.324
	50th - 74th percentile	-.019	.232	-.483	.444
	75th - 99th percentile	.243	.238	-.233	.719
Zscore ASQ 36m	1st - 25th percentile	-.393	.251	-.895	.109
	26th - 49th percentile	.090	.205	-.320	.500
	50th - 74th percentile	.046	.200	-.352	.445
	75th - 99th percentile	.103	.205	-.307	.513
Zscore production 36m	1st - 25th percentile	-.291	.277	-.845	.264
	26th - 49th percentile	-.127	.227	-.579	.326
	50th - 74th percentile	.217	.221	-.224	.658
	75th - 99th percentile	.182	.227	-.270	.635
Zscore sentence Complexity 36m	1st - 25th percentile	-.082	.263	-.607	.443
	26th - 49th percentile	.017	.215	-.411	.446
	50th - 74th percentile	-.025	.209	-.442	.392
	75th - 99th percentile	.073	.215	-.356	.502

Multiple Comparisons

Tukey HSD

Dependent Variable	(I) gesture levels at 18m	(J) gesture levels at 18m	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval Lower Bound	Upper Bound
Zscore comprehension 18m	1st - 25th percentile	26th - 49th percentile	-.7302	.32380	.120	-1.5847	.1243
		50th - 74th percentile	-1.1533*	.32038	.003	-1.9988	-.3079
		75th - 99th percentile	-1.5261*	.32380	.000	-2.3806	-.6716
	26th - 49th percentile	1st - 25th percentile	.7302	.32380	.120	-.1243	1.5847
		50th - 74th percentile	-.4231	.28578	.455	-1.1772	.3311
		75th - 99th percentile	-.7959*	.28962	.038	-1.5602	-.0316
	50th - 74th percentile	1st - 25th percentile	1.1533*	.32038	.003	.3079	1.9988
		26th - 49th percentile	.4231	.28578	.455	-.3311	1.1772
		75th - 99th percentile	-.3728	.28578	.563	-1.1270	.3813
	75th - 99th percentile	1st - 25th percentile	1.5261*	.32380	.000	.6716	2.3806
		26th - 49th percentile	.7959*	.28962	.038	.0316	1.5602
		50th - 74th percentile	.3728	.28578	.563	-.3813	1.1270
Zscore production 18m	1st - 25th percentile	26th - 49th percentile	-.1736	.37047	.966	-1.1512	.8041
		50th - 74th percentile	-.7737	.36655	.161	-1.7410	.1936
		75th - 99th percentile	-.9548	.37047	.058	-1.9325	.0228
	26th - 49th percentile	1st - 25th percentile	.1736	.37047	.966	-.8041	1.1512
		50th - 74th percentile	-.6002	.32697	.267	-1.4630	.2627
		75th - 99th percentile	-.7813	.33136	.096	-1.6557	.0932
	50th - 74th percentile	1st - 25th percentile	.7737	.36655	.161	-.1936	1.7410
		26th - 49th percentile	.6002	.32697	.267	-.2627	1.4630
		75th - 99th percentile	-.1811	.32697	.945	-1.0439	.6818
	75th - 99th percentile	1st - 25th percentile	.9548	.37047	.058	-.0228	1.9325
		26th - 49th percentile	.7813	.33136	.096	-.0932	1.6557
		50th - 74th percentile	.1811	.32697	.945	-.6818	1.0439
Zscore Auditory Comprehension 18m	1st - 25th percentile	26th - 49th percentile	-.8091	.35596	.115	-1.7485	.1302
		50th - 74th percentile	-.9510*	.35219	.043	-1.8805	-.0216
		75th - 99th percentile	-.9791*	.35596	.038	-1.9184	-.0397
	26th - 49th percentile	1st - 25th percentile	.8091	.35596	.115	-.1302	1.7485
		50th - 74th percentile	-.1419	.31416	.969	-.9709	.6872
		75th - 99th percentile	-.1699	.31838	.951	-1.0101	.6703

	50th - 74th percentile	1st - 25th percentile	.9510*	.35219	.043	.0216	1.8805
		26th - 49th percentile	.1419	.31416	.969	-.6872	.9709
		75th - 99th percentile	-.0281	.31416	1.000	-.8571	.8010
	75th - 99th percentile	1st - 25th percentile	.9791*	.35596	.038	.0397	1.9184
		26th - 49th percentile	.1699	.31838	.951	-.6703	1.0101
		50th - 74th percentile	.0281	.31416	1.000	-.8010	.8571
Zscore Expressive Communication 18m	1st - 25th percentile	26th - 49th percentile	-.1915	.36161	.952	-1.1458	.7628
		50th - 74th percentile	-.3408	.35779	.777	-1.2850	.6033
		75th - 99th percentile	-.4964	.36161	.521	-1.4507	.4578
	26th - 49th percentile	1st - 25th percentile	.1915	.36161	.952	-.7628	1.1458
		50th - 74th percentile	-.1493	.31915	.966	-.9916	.6929
		75th - 99th percentile	-.3049	.32344	.782	-1.1585	.5486
	50th - 74th percentile	1st - 25th percentile	.3408	.35779	.777	-.6033	1.2850
		26th - 49th percentile	.1493	.31915	.966	-.6929	.9916
		75th - 99th percentile	-.1556	.31915	.962	-.9978	.6866
	75th - 99th percentile	1st - 25th percentile	.4964	.36161	.521	-.4578	1.4507
		26th - 49th percentile	.3049	.32344	.782	-.5486	1.1585
		50th - 74th percentile	.1556	.31915	.962	-.6866	.9978
Zscore comprehension 24m	1st - 25th percentile	26th - 49th percentile	-.3810	.35988	.716	-1.3307	.5687
		50th - 74th percentile	-.7958	.35607	.125	-1.7354	.1438
		75th - 99th percentile	-.8814	.35988	.078	-1.8311	.0683
	26th - 49th percentile	1st - 25th percentile	.3810	.35988	.716	-.5687	1.3307
		50th - 74th percentile	-.4148	.31762	.563	-1.2530	.4234
		75th - 99th percentile	-.5004	.32188	.412	-1.3498	.3490
	50th - 74th percentile	1st - 25th percentile	.7958	.35607	.125	-.1438	1.7354
		26th - 49th percentile	.4148	.31762	.563	-.4234	1.2530
		75th - 99th percentile	-.0856	.31762	.993	-.9238	.7526
	75th - 99th percentile	1st - 25th percentile	.8814	.35988	.078	-.0683	1.8311
		26th - 49th percentile	.5004	.32188	.412	-.3490	1.3498
		50th - 74th percentile	.0856	.31762	.993	-.7526	.9238
Zscore production 24m	1st - 25th percentile	26th - 49th percentile	-.2449	.38485	.920	-1.2605	.7707
		50th - 74th percentile	-.5996	.38078	.400	-1.6045	.4053
		75th - 99th percentile	-.7705	.38485	.198	-1.7861	.2451
	26th - 49th percentile	1st - 25th percentile	.2449	.38485	.920	-.7707	1.2605
		50th - 74th percentile	-.3547	.33966	.724	-1.2510	.5417
		75th - 99th percentile	-.5256	.34422	.428	-1.4340	.3828
	50th - 74th percentile	1st - 25th percentile	.5996	.38078	.400	-.4053	1.6045
		26th - 49th percentile	.3547	.33966	.724	-.5417	1.2510
		75th - 99th percentile	-.1709	.33966	.958	-1.0673	.7255
	75th - 99th percentile	1st - 25th percentile	.7705	.38485	.198	-.2451	1.7861
		26th - 49th percentile	.5256	.34422	.428	-.3828	1.4340

		50th - 74th percentile	.1709	.33966	.958	-.7255	1.0673
Zscore sentence Complexity 24m	1st - 25th percentile	26th - 49th percentile	-.1210	.36792	.988	-1.0919	.8499
		50th - 74th percentile	-.5659	.36402	.412	-1.5266	.3947
		75th - 99th percentile	-.4897	.36792	.547	-1.4606	.4812
	26th - 49th percentile	1st - 25th percentile	.1210	.36792	.988	-.8499	1.0919
		50th - 74th percentile	-.4449	.32472	.523	-1.3018	.4120
		75th - 99th percentile	-.3687	.32908	.678	-1.2371	.4998
	50th - 74th percentile	1st - 25th percentile	.5659	.36402	.412	-.3947	1.5266
		26th - 49th percentile	.4449	.32472	.523	-.4120	1.3018
		75th - 99th percentile	.0763	.32472	.995	-.7807	.9332
	75th - 99th percentile	1st - 25th percentile	.4897	.36792	.547	-.4812	1.4606
		26th - 49th percentile	.3687	.32908	.678	-.4998	1.2371
		50th - 74th percentile	-.0763	.32472	.995	-.9332	.7807
Zscore Auditory Comprehension 24m	1st - 25th percentile	26th - 49th percentile	-.5635	.35987	.405	-1.5132	.3862
		50th - 74th percentile	-.6012	.35606	.338	-1.5409	.3384
		75th - 99th percentile	-1.0026*	.35987	.035	-1.9522	-.0529
	26th - 49th percentile	1st - 25th percentile	.5635	.35987	.405	-.3862	1.5132
		50th - 74th percentile	-.0377	.31761	.999	-.8759	.8004
		75th - 99th percentile	-.4391	.32188	.526	-1.2885	.4103
	50th - 74th percentile	1st - 25th percentile	.6012	.35606	.338	-.3384	1.5409
		26th - 49th percentile	.0377	.31761	.999	-.8004	.8759
		75th - 99th percentile	-.4013	.31761	.589	-1.2395	.4368
	75th - 99th percentile	1st - 25th percentile	1.0026*	.35987	.035	.0529	1.9522
		26th - 49th percentile	.4391	.32188	.526	-.4103	1.2885
		50th - 74th percentile	.4013	.31761	.589	-.4368	1.2395
Zscore Expressive Communication 24m	1st - 25th percentile	26th - 49th percentile	.0465	.37673	.999	-.9476	1.0407
		50th - 74th percentile	-.0863	.37274	.996	-1.0700	.8973
		75th - 99th percentile	-.3489	.37673	.791	-1.3431	.6452
	26th - 49th percentile	1st - 25th percentile	-.0465	.37673	.999	-1.0407	.9476
		50th - 74th percentile	-.1328	.33249	.978	-1.0103	.7446
		75th - 99th percentile	-.3955	.33696	.646	-1.2847	.4938
	50th - 74th percentile	1st - 25th percentile	.0863	.37274	.996	-.8973	1.0700
		26th - 49th percentile	.1328	.33249	.978	-.7446	1.0103
		75th - 99th percentile	-.2626	.33249	.859	-1.1400	.6148
	75th - 99th percentile	1st - 25th percentile	.3489	.37673	.791	-.6452	1.3431
		26th - 49th percentile	.3955	.33696	.646	-.4938	1.2847
		50th - 74th percentile	.2626	.33249	.859	-.6148	1.1400
Zscore ASQ 36m	1st - 25th percentile	26th - 49th percentile	-.4831	.32427	.450	-1.3388	.3727
		50th - 74th percentile	-.4397	.32084	.522	-1.2864	.4070
		75th - 99th percentile	-.4963	.32427	.426	-1.3520	.3594
	26th - 49th percentile	1st - 25th percentile	.4831	.32427	.450	-.3727	1.3388
		50th - 74th percentile	.0434	.28620	.999	-.7119	.7986

		75th - 99th percentile	-.0132	.29004	1.000	-.7786	.7522
	50th - 74th percentile	1st - 25th percentile	.4397	.32084	.522	-.4070	1.2864
		26th - 49th percentile	-.0434	.28620	.999	-.7986	.7119
		75th - 99th percentile	-.0566	.28620	.997	-.8119	.6986
	75th - 99th percentile	1st - 25th percentile	.4963	.32427	.426	-.3594	1.3520
		26th - 49th percentile	.0132	.29004	1.000	-.7522	.7786
		50th - 74th percentile	.0566	.28620	.997	-.6986	.8119
Zscore production 36m	1st - 25th percentile	26th - 49th percentile	-.1641	.35821	.968	-1.1094	.7812
		50th - 74th percentile	-.5077	.35442	.484	-1.4430	.4276
		75th - 99th percentile	-.4733	.35821	.553	-1.4186	.4720
	26th - 49th percentile	1st - 25th percentile	.1641	.35821	.968	-.7812	1.1094
		50th - 74th percentile	-.3436	.31615	.699	-1.1779	.4907
		75th - 99th percentile	-.3091	.32039	.770	-1.1546	.5364
	50th - 74th percentile	1st - 25th percentile	.5077	.35442	.484	-.4276	1.4430
		26th - 49th percentile	.3436	.31615	.699	-.4907	1.1779
		75th - 99th percentile	.0344	.31615	1.000	-.7999	.8687
	75th - 99th percentile	1st - 25th percentile	.4733	.35821	.553	-.4720	1.4186
		26th - 49th percentile	.3091	.32039	.770	-.5364	1.1546
		50th - 74th percentile	-.0344	.31615	1.000	-.8687	.7999
Zscore sentence Complexity 36m	1st - 25th percentile	26th - 49th percentile	-.0997	.33919	.991	-.9948	.7954
		50th - 74th percentile	-.0575	.33560	.998	-.9431	.8282
		75th - 99th percentile	-.1553	.33919	.968	-1.0504	.7398
	26th - 49th percentile	1st - 25th percentile	.0997	.33919	.991	-.7954	.9948
		50th - 74th percentile	.0422	.29936	.999	-.7478	.8322
		75th - 99th percentile	-.0556	.30338	.998	-.8562	.7450
	50th - 74th percentile	1st - 25th percentile	.0575	.33560	.998	-.8282	.9431
		26th - 49th percentile	-.0422	.29936	.999	-.8322	.7478
		75th - 99th percentile	-.0978	.29936	.988	-.8878	.6922
	75th - 99th percentile	1st - 25th percentile	.1553	.33919	.968	-.7398	1.0504
		26th - 49th percentile	.0556	.30338	.998	-.7450	.8562
		50th - 74th percentile	.0978	.29936	.988	-.6922	.8878

Based on observed means.

The error term is Mean Square(Error) = .828.

*. The mean difference is significant at the .05 level.

Appendix 14 - Difference between hours of sleep at 12 and 18 months

Descriptive Statistics

	N	Mean	Std. Deviation	Minimum	Maximum
sleep12	72	13.0000	1.31084	9.00	15.00
sleep18	71	12.7183	1.09783	9.00	15.00

Ranks

		N	Mean Rank	Sum of Ranks
sleep18 - sleep12	Negative Ranks	26[a]	18.15	472.00
	Positive Ranks	10[b]	19.40	194.00
	Ties	33[c]		
	Total	69		

a. sleep18 < sleep12

b. sleep18 > sleep12

c. sleep18 = sleep12

Test Statistics[a]

	sleep18 - sleep12
Z	-2.305[b]
Asymp. Sig. (2-tailed)	.021

a. Wilcoxon Signed Ranks Test

b. Based on positive ranks.

Appendix 15 – Moderation analyses

It was investigated if cognitive skills at 18 months had a moderating effect on the relationship between PLS scores at 18 and 24 months, but the interaction was not significant (p = .796). The moderating effect for cognitive skills at 24 months between CDI scores and later language skills was also investigated but were not significant, see Table 1.

Table 1. *Moderating effects of cognitive skills at 24m on the relationship between the following communication abilities*

Predictor	Outcome	Sig. (p value)
Gestures 12m	Comprehension 18m	.755
Gestures 12m	PLS Total Language 18m	.094
Gestures 12m	comprehension 24m	.292
Gestures 12m	PLS Total Language 24m	.829
Gestures 12m	Sentence Complexity 36m	.874
Comprehension 18m	Comprehension 24m	.723
Comprehension 18m	PLS Total Language 24m	.766
Comprehension 18m	Sentence Complexity 36m	.334

The next step was to test for moderating effects of motor skills at 18 months between the PLS scores at 18 and 24 months, but motor skills at 18 months was not a significant moderator (p = .791). Furthermore, motor skills at 24 months was not significant moderator, see Table 2.

Table 2. *Moderating effects of motor skills at 24 months on the relationship between the following communication abilities*

Predictor	Outcome	sig. (p value)
Comprehension 18m	PLS 24m	.300
Phrases Understood 18m	PLS 24m	.205

Appendix 16 - ROC curve analyses

UK-CDI scores at 18m predicting language delay on the PLS at 24m

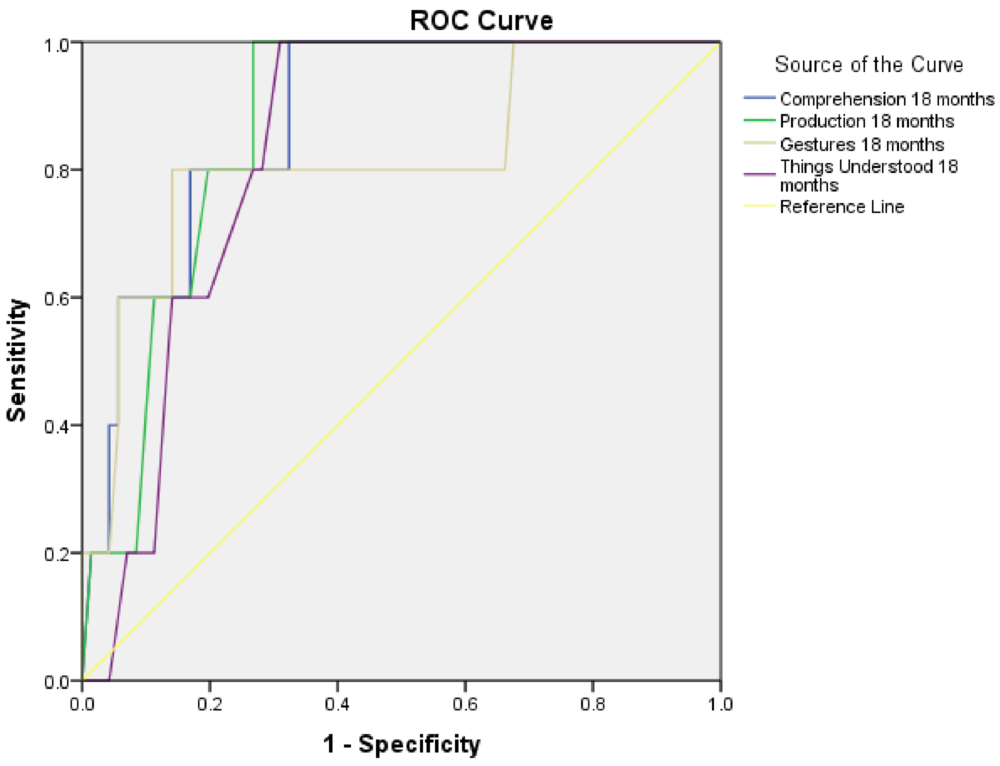
Case Processing Summary

Delay on PLS at 24 months[a]	Valid N (listwise)
Positive[b]	5
Negative	71
Missing	2

Smaller values of the test result variable(s) indicate stronger evidence for a positive actual state.

a. The test result variable(s): Phrases Understood 18 months has at least one tie between the positive actual state group and the negative actual state group.

b. The positive actual state is 1.00.



Diagonal segments are produced by ties.

Area Under the Curve

Test Result Variable(s)	Area	Std. Error[a]	Asymptotic Sig.[b]	Asymptotic 95% Confidence Interval	
				Lower Bound	Upper Bound
Comprehension 18 months	.880	.059	.005	.765	.995
Production 18 months	.869	.051	.006	.769	.969
Gestures 18 months	.817	.113	.018	.595	1.000
Phrases Understood 18 months	.832	.054	.013	.727	.938

The test result variable(s): Comprehension 18 months, Production 18 months, Gestures 18 months , Phrases Understood 18 months has at least one tie between the positive actual state group and the negative actual state group. Statistics may be biased.

a. Under the nonparametric assumption

b. Null hypothesis: true area = 0.5

Coordinates of the Curve

Test Result Variable(s)	Positive if Less Than or Equal To[a]	Sensitivity	1 - Specificity
Comprehension 18 months	72.0000	.000	.000
	74.5000	.200	.014
	77.0000	.200	.028
	82.0000	.200	.042
	89.0000	.400	.042
	94.5000	.400	.056
	97.5000	.600	.056
	101.5000	.600	.070
	109.0000	.600	.085
	113.5000	.600	.099
	117.0000	.600	.113
	121.0000	.600	.127
	124.5000	.600	.141
	129.0000	.600	.155
	134.0000	.600	.169
	138.5000	.800	.169
	142.0000	.800	.183
	144.5000	.800	.197
	146.5000	.800	.211

	148.5000	.800	.225
	150.5000	.800	.239
	154.0000	.800	.268
	164.5000	.800	.282
	173.5000	.800	.310
	176.0000	.800	.324
	181.0000	1.000	.324
	185.0000	1.000	.338
	187.0000	1.000	.352
	189.0000	1.000	.366
	197.0000	1.000	.394
	205.5000	1.000	.408
	208.0000	1.000	.423
	209.5000	1.000	.437
	210.5000	1.000	.451
	213.5000	1.000	.465
	218.5000	1.000	.479
	222.5000	1.000	.493
	226.0000	1.000	.507
	228.5000	1.000	.521
	229.5000	1.000	.549
	230.5000	1.000	.563
	232.5000	1.000	.577
	236.0000	1.000	.592
	238.5000	1.000	.606
	239.5000	1.000	.634
	241.0000	1.000	.648
	244.5000	1.000	.662
	253.0000	1.000	.676
	262.0000	1.000	.690
	266.0000	1.000	.704
	268.0000	1.000	.718
	272.5000	1.000	.732
	281.0000	1.000	.746
	287.5000	1.000	.761
	292.5000	1.000	.775
	299.0000	1.000	.789
	302.5000	1.000	.803
	305.5000	1.000	.817
	309.0000	1.000	.831

	313.5000	1.000	.845
	321.5000	1.000	.873
	332.0000	1.000	.887
	342.0000	1.000	.901
	347.5000	1.000	.915
	355.5000	1.000	.930
	371.5000	1.000	.944
	382.0000	1.000	.958
	385.5000	1.000	.972
	391.5000	1.000	.986
	396.0000	1.000	1.000
Production 18 months	1.0000	.000	.000
	3.0000	.200	.014
	5.5000	.200	.056
	7.5000	.200	.085
	8.5000	.400	.099
	10.0000	.600	.113
	11.5000	.600	.127
	13.0000	.600	.141
	15.0000	.600	.169
	17.0000	.800	.197
	18.5000	.800	.211
	19.5000	.800	.225
	20.5000	.800	.254
	21.5000	.800	.268
	22.5000	1.000	.268
	23.5000	1.000	.282
	25.0000	1.000	.296
	26.5000	1.000	.310
	27.5000	1.000	.324
	28.5000	1.000	.338
	29.5000	1.000	.352
	30.5000	1.000	.366
	31.5000	1.000	.394
	33.0000	1.000	.408
	35.0000	1.000	.423
	37.5000	1.000	.451
	40.0000	1.000	.479
	42.5000	1.000	.493
	45.5000	1.000	.507
	47.5000	1.000	.535
	50.0000	1.000	.549
	52.5000	1.000	.563
	55.5000	1.000	.577

	59.0000	1.000	.592
	63.0000	1.000	.620
	67.0000	1.000	.648
	71.5000	1.000	.662
	75.5000	1.000	.676
	78.0000	1.000	.690
	83.0000	1.000	.704
	87.5000	1.000	.718
	95.0000	1.000	.732
	101.5000	1.000	.746
	102.5000	1.000	.761
	105.5000	1.000	.803
	111.0000	1.000	.817
	115.0000	1.000	.831
	119.0000	1.000	.845
	129.0000	1.000	.859
	136.5000	1.000	.873
	138.5000	1.000	.887
	147.0000	1.000	.901
	164.0000	1.000	.915
	181.5000	1.000	.930
	199.0000	1.000	.944
	236.5000	1.000	.958
	272.5000	1.000	.972
	298.5000	1.000	.986
	317.0000	1.000	1.000
Gestures 18 months	19.0000	.000	.000
	21.2500	.200	.000
	24.2500	.200	.014
	28.0000	.200	.028
	30.5000	.200	.042
	31.5000	.400	.056
	32.2500	.600	.056
	33.0000	.600	.070
	34.0000	.600	.085
	35.2500	.600	.099
	36.7500	.600	.113
	37.7500	.600	.127
	38.2500	.600	.141
	38.7500	.800	.141
	39.5000	.800	.183
	40.2500	.800	.225
	40.7500	.800	.239
	41.2500	.800	.254
	41.7500	.800	.268

	42.2500	.800	.324
	42.7500	.800	.352
	43.2500	.800	.423
	43.7500	.800	.465
	44.2500	.800	.521
	44.7500	.800	.549
	45.2500	.800	.563
	45.7500	.800	.577
	46.2500	.800	.592
	46.7500	.800	.606
	47.5000	.800	.634
	48.2500	.800	.662
	48.7500	1.000	.676
	49.2500	1.000	.704
	50.0000	1.000	.718
	50.7500	1.000	.746
	51.2500	1.000	.761
	52.0000	1.000	.817
	52.7500	1.000	.831
	53.5000	1.000	.845
	54.5000	1.000	.859
	55.2500	1.000	.873
	55.7500	1.000	.901
	56.2500	1.000	.915
	57.0000	1.000	.930
	58.0000	1.000	.944
	59.0000	1.000	.958
	59.7500	1.000	.972
	61.0000	1.000	.986
	63.0000	1.000	1.000
Phrases Understood 18 months	10.0000	.000	.000
	12.0000	.000	.028
	13.5000	.000	.042
	14.5000	.200	.070
	15.5000	.200	.085
	16.5000	.200	.113
	17.5000	.600	.141
	18.5000	.600	.197
	19.5000	.800	.268
	20.5000	.800	.282
	21.5000	1.000	.310
	22.5000	1.000	.380
	23.5000	1.000	.493

	24.5000	1.000	.592
	25.5000	1.000	.648
	26.5000	1.000	.732
	27.5000	1.000	.873
	29.0000	1.000	1.000

The test result variable(s): Comprehension 18 months, Production 18 months, Gestures 18 months, Phrases Understood 18 months has at least one tie between the positive actual state group and the negative actual state group.
a. The smallest cutoff value is the minimum observed test value minus 1, and the largest cutoff value is the maximum observed test value plus 1. All the other cutoff values are the averages of two consecutive ordered observed test values.

Appendix 17 - Logistic regressions: predicting language delay at 24m and 36m from Production 12m scores after controlling for age at 12m

Delay criterion: Rescorla's criterion at 24 months

Case Processing Summary

Unweighted Cases[a]		N	Percent
Selected Cases	Included in Analysis	78	100.0
	Missing Cases	0	.0
	Total	78	100.0
Unselected Cases		0	.0
Total		78	100.0

a. If weight is in effect, see classification table for the total number of cases.

Dependent Variable Encoding

Original Value	Internal Value
.00	0
1.00	1

Block 0: Beginning Block

Classification Table[a,b]

	Observed		Predicted		
			Rescorla criterion		Percentage Correct
			.00	1.00	
Step 0	Rescorla criterion	.00	67	0	100.0
		1.00	11	0	.0
	Overall Percentage				85.9

a. Constant is included in the model.

b. The cut value is .500

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 0	Constant	-1.807	.325	30.845	1	.000	.164

Variables not in the Equation

			Score	df	Sig.
Step 0	Variables	age12	.021	1	.884
	Overall Statistics		.021	1	.884

Block 1: Method = Enter

Omnibus Tests of Model Coefficients

		Chi-square	df	Sig.
Step 1	Step	.021	1	.885
	Block	.021	1	.885
	Model	.021	1	.885

Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	63.443[a]	.000	.000

a. Estimation terminated at iteration number 5 because parameter estimates changed by less than .001.

Classification Table[a]

	Observed		Predicted		
			Rescorla criterion		Percentage Correct
			.00	1.00	
Step 1	Rescorla criterion	.00	67	0	100.0
		1.00	11	0	.0
	Overall Percentage				85.9

a. The cut value is .500

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1[a]	age12	.040	.271	.021	1	.884	1.040
	Constant	-2.270	3.196	.504	1	.478	.103

a. Variable(s) entered on step 1: age12.

Block 2: Method = Enter

Omnibus Tests of Model Coefficients

		Chi-square	df	Sig.
Step 1	Step	3.772	1	.052
	Block	3.772	1	.052
	Model	3.793	2	.150

Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	59.671[a]	.047	.085

a. Estimation terminated at iteration number 5 because parameter estimates changed by less than .001.

Classification Table[a]

	Observed		Predicted		
			Rescorla criterion		Percentage Correct
			.00	1.00	
Step 1	Rescorla criterion	.00	67	0	100.0
		1.00	11	0	.0
	Overall Percentage				85.9

a. The cut value is .500

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1[a]	age12	.128	.303	.179	1	.672	1.137
	Zprod12t	-.711	.394	3.253	1	.071	.491
	Constant	-3.443	3.573	.928	1	.335	.032

a. Variable(s) entered on step 1: Zprod12t.

Delay criterion: Production below 20th percentile or no word combinations at 24 months

Case Processing Summary

Unweighted Cases[a]		N	Percent
Selected Cases	Included in Analysis	78	100.0
	Missing Cases	0	.0
	Total	78	100.0
Unselected Cases		0	.0
Total		78	100.0

a. If weight is in effect, see classification table for the total number of cases.

Dependent Variable Encoding

Original Value	Internal Value
.00	0
1.00	1

Block 0: Beginning Block

Classification Table[a,b]

	Observed		Predicted		
			below 20th percentile or no word combinations		Percentage Correct
			.00	1.00	
Step 0	below 20th percentile or no word combinations	.00	61	0	100.0
		1.00	17	0	.0
	Overall Percentage				78.2

a. Constant is included in the model.

b. The cut value is .500

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 0	Constant	-1.278	.274	21.703	1	.000	.279

Variables not in the Equation

			Score	df	Sig.
Step 0	Variables	age12	.017	1	.897
	Overall Statistics		.017	1	.897

Block 1: Method = Enter

Omnibus Tests of Model Coefficients

		Chi-square	df	Sig.
Step 1	Step	.017	1	.897
	Block	.017	1	.897
	Model	.017	1	.897

Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	81.774[a]	.000	.000

a. Estimation terminated at iteration number 4 because parameter estimates changed by less than .001.

Classification Table[a]

	Observed		Predicted		
			below 20th percentile or no word combinations		Percentage Correct
			.00	1.00	
Step 1	below 20th percentile or no word combinations	.00	61	0	100.0
		1.00	17	0	.0
	Overall Percentage				78.2

a. The cut value is .500

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1[a]	age12	.061	.275	.049	1	.824	1.063
	Zprod12t	-.902	.357	6.398	1	.011	.406
	Constant	-2.164	3.214	.453	1	.501	.115

a. Variable(s) entered on step 1: Zprod12t.

Delay criterion: PLS at 24 months

Case Processing Summary

Unweighted Cases[a]		N	Percent
Selected Cases	Included in Analysis	78	100.0
	Missing Cases	0	.0
	Total	78	100.0
Unselected Cases		0	.0
Total		78	100.0

a. If weight is in effect, see classification table for the total number of cases.

Dependent Variable Encoding

Original Value	Internal Value
.00	0
1.00	1

Block 0: Beginning Block

Classification Table[a,b]

	Observed		Predicted		
			Delay on PLS at 24 months		Percentage Correct
			.00	1.00	
Step 0	Delay on PLS at 24 months	.00	73	0	100.0
		1.00	5	0	.0
	Overall Percentage				93.6

a. Constant is included in the model.

b. The cut value is .500

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 0	Constant	-2.681	.462	33.636	1	.000	.068

Variables not in the Equation

			Score	df	Sig.
Step 0	Variables	age12	.025	1	.875
	Overall Statistics		.025	1	.875

Block 1: Method = Enter

Omnibus Tests of Model Coefficients

		Chi-square	df	Sig.
Step 1	Step	.025	1	.873
	Block	.025	1	.873
	Model	.025	1	.873

Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	37.120[a]	.000	.001

a. Estimation terminated at iteration number 6 because parameter estimates changed by less than .001.

Classification Table[a]

Observed			Predicted: Delay on PLS at 24 months .00	Predicted: Delay on PLS at 24 months 1.00	Predicted: Percentage Correct
Step 1	Delay on PLS at 24 months	.00	73	0	100.0
		1.00	5	0	.0
	Overall Percentage				93.6

a. The cut value is .500

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1[a]	age12	-.065	.414	.025	1	.875	.937
	Constant	-1.924	4.832	.159	1	.690	.146

a. Variable(s) entered on step 1: age12.

Block 2: Method = Enter

Omnibus Tests of Model Coefficients

		Chi-square	df	Sig.
Step 1	Step	3.751	1	.053
	Block	3.751	1	.053
	Model	3.776	2	.151

Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	33.369[a]	.047	.125

a. Estimation terminated at iteration number 7 because parameter estimates changed by less than .001.

Classification Table[a]

	Observed		Predicted		
			Delay on PLS at 24 months		Percentage Correct
			.00	1.00	
Step 1	Delay on PLS at 24 months	.00	73	0	100.0
		1.00	5	0	.0
	Overall Percentage				93.6

a. The cut value is .500

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1[a]	age12	-.031	.463	.005	1	.946	.969
	Zprod12t	-1.130	.671	2.836	1	.092	.323
	Constant	-2.737	5.352	.262	1	.609	.065

a. Variable(s) entered on step 1: Zprod12t.

Delay: below the 19th percentile for Production on the 3-year parent report language

Case Processing Summary

Unweighted Cases[a]		N	Percent
Selected Cases	Included in Analysis	82	100.0
	Missing Cases	0	.0
	Total	82	100.0
Unselected Cases		0	.0
Total		82	100.0

a. If weight is in effect, see classification table for the total number of cases.

Dependent Variable Encoding

Original Value	Internal Value
.00	0
1.00	1

Block 0: Beginning Block

Classification Table[a,b]

Observed			Predicted: delay 19th percentile .00	Predicted: delay 19th percentile 1.00	Percentage Correct
Step 0	delay 19th percentile	.00	67	0	100.0
		1.00	15	0	.0
	Overall Percentage				81.7

a. Constant is included in the model.

b. The cut value is .500

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 0	Constant	-1.497	.286	27.453	1	.000	.224

Variables not in the Equation

			Score	df	Sig.
Step 0	Variables	age12	.332	1	.564
	Overall Statistics		.332	1	.564

Block 1: Method = Enter

Omnibus Tests of Model Coefficients

		Chi-square	df	Sig.
Step 1	Step	.319	1	.572
	Block	.319	1	.572
	Model	.319	1	.572

Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	77.713[a]	.004	.006

a. Estimation terminated at iteration number 4 because parameter estimates changed by less than .001.

Classification Table[a]

	Observed		Predicted		
			delay 19th percentile		Percentage Correct
			.00	1.00	
Step 1	delay 19th percentile	.00	67	0	100.0
		1.00	15	0	.0
	Overall Percentage				81.7

a. The cut value is .500

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1[a]	age12	.131	.228	.330	1	.566	1.140
	Constant	-3.036	2.708	1.256	1	.262	.048

a. Variable(s) entered on step 1: age12.

Block 2: Method = Enter

Omnibus Tests of Model Coefficients

		Chi-square	df	Sig.
Step 1	Step	3.602	1	.058
	Block	3.602	1	.058
	Model	3.920	2	.141

Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	74.111[a]	.047	.076

a. Estimation terminated at iteration number 5 because parameter estimates changed by less than .001.

Classification Table[a]

	Observed		Predicted		
			delay 19th percentile		Percentage Correct
			.00	1.00	
Step 1	delay 19th percentile	.00	67	0	100.0
		1.00	15	0	.0
	Overall Percentage				81.7

a. The cut value is .500

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1[a]	age12	.253	.254	.995	1	.319	1.288
	Zprod12t	-.589	.327	3.257	1	.071	.555
	Constant	-4.529	3.027	2.238	1	.135	.011

a. Variable(s) entered on step 1: Zprod12t.

Delay: ASQ communication below 2 SD at 36m

Case Processing Summary

Unweighted Cases[a]		N	Percent
Selected Cases	Included in Analysis	82	100.0
	Missing Cases	0	.0
	Total	82	100.0
Unselected Cases		0	.0
Total		82	100.0

a. If weight is in effect, see classification table for the total number of cases.

Dependent Variable Encoding

Original Value	Internal Value
.00	0
1.00	1

Block 0: Beginning Block

Classification Table[a,b]

	Observed		Predicted		
			severe delay on ASQ		Percentage Correct
			.00	1.00	
Step 0	severe delay on ASQ	.00	78	0	100.0
		1.00	4	0	.0
	Overall Percentage				95.1

a. Constant is included in the model.

b. The cut value is .500

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 0	Constant	-2.970	.513	33.572	1	.000	.051

Variables not in the Equation

			Score	df	Sig.
Step 0	Variables	age12	5.012	1	.025
	Overall Statistics		5.012	1	.025

Block 1: Method = Enter

Omnibus Tests of Model Coefficients

		Chi-square	df	Sig.
Step 1	Step	3.752	1	.053
	Block	3.752	1	.053
	Model	3.752	1	.053

Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	28.213[a]	.045	.139

a. Estimation terminated at iteration number 6 because parameter estimates changed by less than .001.

Classification Table[a]

	Observed		Predicted		
			severe delay on ASQ		Percentage Correct
			.00	1.00	
Step 1	severe delay on ASQ	.00	78	0	100.0
		1.00	4	0	.0
	Overall Percentage				95.1

a. The cut value is .500

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1[a]	age12	1.005	.432	5.417	1	.020	2.733
	Zprod12t	-.938	.557	2.836	1	.092	.392
	Constant	-15.488	5.645	7.528	1	.006	.000

a. Variable(s) entered on step 1: Zprod12t.

Delay: ASQ communication below 1 SD at 36m

Case Processing Summary

Unweighted Cases[a]		N	Percent
Selected Cases	Included in Analysis	82	100.0
	Missing Cases	0	.0
	Total	82	100.0
Unselected Cases		0	.0
Total		82	100.0

a. If weight is in effect, see classification table for the total number of cases.

Dependent Variable Encoding

Original Value	Internal Value
.00	0
1.00	1

Block 0: Beginning Block

Classification Table[a,b]

	Observed		Predicted		
			ASQcommdelay1		Percentage Correct
			.00	1.00	
Step 0	ASQcommdelay1	.00	72	0	100.0
		1.00	10	0	.0
	Overall Percentage				87.8

a. Constant is included in the model.

b. The cut value is .500

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 0	Constant	-1.974	.337	34.218	1	.000	.139

Variables not in the Equation

			Score	df	Sig.
Step 0	Variables	age12	.696	1	.404
	Overall Statistics		.696	1	.404

Block 1: Method = Enter

Omnibus Tests of Model Coefficients

		Chi-square	df	Sig.
Step 1	Step	.641	1	.424
	Block	.641	1	.424
	Model	.641	1	.424

Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	60.170[a]	.008	.015

a. Estimation terminated at iteration number 5 because parameter estimates changed by less than .001.

Classification Table[a]

	Observed		Predicted		
			ASQcommdelay1		Percentage Correct
			.00	1.00	
Step 1	ASQcommdelay1	.00	72	0	100.0
		1.00	10	0	.0
	Overall Percentage				87.8

a. The cut value is .500

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1[a]	age12	.211	.256	.682	1	.409	1.235
	Constant	-4.472	3.075	2.115	1	.146	.011

a. Variable(s) entered on step 1: age12.

Block 2: Method = Enter

Omnibus Tests of Model Coefficients

		Chi-square	df	Sig.
Step 1	Step	1.150	1	.284
	Block	1.150	1	.284
	Model	1.791	2	.409

Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	59.020[a]	.022	.041

a. Estimation terminated at iteration number 5 because parameter estimates changed by less than .001.

Classification Table[a]

	Observed		Predicted		
			ASQcommdelay1		Percentage Correct
			.00	1.00	
Step 1	ASQcommdelay1	.00	72	0	100.0
		1.00	10	0	.0
	Overall Percentage				87.8

a. The cut value is .500

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1[a]	age12	.303	.279	1.173	1	.279	1.353
	Zprod12t	-.373	.356	1.098	1	.295	.689
	Constant	-5.568	3.367	2.735	1	.098	.004

a. Variable(s) entered on step 1: Zprod12t.